

Result No.	Score	Query Match	Length	DB ID	Description
1	3150	100.0	599	22 AAE04784	Arabidopsis thalia
2	3150	100.0	599	23 ABB92311	Herbicidally activ
3	2280.5	72.4	605	22 AAE04789	Lycopersicon escul
4	2280.5	72.4	605	22 AAE72308	Neoxanthin cleavag
5	2168.5	68.8	612	22 AAE04787	Vigna unguiculata
6	2044.5	64.9	589	23 ABB91182	Herbicidally activ
7	1991	63.2	583	22 AAE04782	Arabidopsis thalia
8	1930	61.3	604	22 AAE04788	Zea mays. aeoxanthi
9	1930	61.3	604	22 AAE72309	Neoxanthin cleavag
10	1663.5	52.8	577	22 AAE04786	Arabidopsis thalia

PT producing transgenic plants with improved or decreased stress tolerance
 PS
 XX
 XX
 XX Claim 3; Fig 9; 101pp; English.
 CC The invention relates to neoxanthin cleavage enzymes and their
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 CC plant when expressed in a plant cell. The invention also relates to
 CC methods for increasing or decreasing stress tolerance in a plant by
 CC introducing the DNA into the plant, and a transgenic plant into which a
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid
 CC land can be improved by growing transformant weed for several years and
 CC then removing the weed by specifically lowering stress tolerance in the
 CC weed by inducing an inducible promoter. The present sequence is
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 protein.
 CC The AtNCED3 cDNA is obtained from an Arabidopsis plant-derived cDNA
 CC library using a cDNA of the cPRD65 (CovPea Responsive to Dehydration)
 CC gene isolated from cowpea plant as a probe.

XX
 SQ Sequence 599 AA;

Query Match 100.0%; Score 3150; DB 22; Length 599;
 Best Local Similarity 100.0%; Pred. No. 2.6e-310;
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASFTATAAAGSRWLGNGHNTOPPLSSOSSDLSYCSSLPMASRVTRKLVNSSLHTPPAL 60
 DB 1 MASFTATAAAGSRWLGNGHNTOPPLSSOSSDLSYCSSLPMASRVTRKLVNSSLHTPPAL 60
 QY 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 DB 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 QY 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTHGFFDGDGMVHAV 180
 DB 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTHGFFDGDGMVHAV 180
 QY 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPKAIGELHGHGTGIARLMLFYARAAGIVDP 240
 DB 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPKAIGELHGHGTGIARLMLFYARAAGIVDP 240
 QY 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTVGRFDFDGLQESTMTIAHPKV 300
 DB 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTVGRFDFDGLQESTMTIAHPKV 300
 QY 301 DPESGELFALSYDVVSKPKYKFRSPDGTSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 DB 301 DPESGELFALSYDVVSKPKYKFRSPDGTSPDGTSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 QY 361 QVVFKLPEMIRGSPVVDKKNVAREGILDKYAEDSSNLIKWDADPCFCFHLNAAWEPE 420
 DB 361 QVVFKLPEMIRGSPVVDKKNVAREGILDKYAEDSSNLIKWDADPCFCFHLNAAWEPE 420
 QY 421 TDEVVVGSCMTPPDSIFNEDENLKSVLSEIRLNKLTGESIRRPISNEDQOVNLEAGM 480
 DB 421 TDEVVVGSCMTPPDSIFNEDENLKSVLSEIRLNKLTGESIRRPISNEDQOVNLEAGM 480
 QY 481 VNRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGE 540
 DB 481 VNRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGE 540
 QY 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSRPVPGFHGFTFGAODLAKOVV 599
 DB 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSRPVPGFHGFTFGAODLAKOVV 599

RESULT 2
 ABB92311
 ID ABB92311 standard; Protein; 599 AA.

XX
 AC ABB92311;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 DE Herbicidally active polypeptide SEQ ID NO 1522.
 XX
 KW Herbicidal; plant; agriculture; herbicide.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (PABB) BAYER AG.
 XX
 PI Tietjen K, Weidler M;
 XX
 DR WPI; 2002-269010/31.
 XX
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 FS Claim 5; SEQ ID NO 1522; 261pp + Sequence Listing; English.
 CC
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.
 XX
 SQ Sequence 599 AA;
 Query Match 100.0%; Score 3150; DB 23; Length 599;
 Best Local Similarity 100.0%; Pred. No. 2.6e-310;
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASFTATAAAGSRWLGNGHNTOPPLSSOSSDLSYCSSLPMASRVTRKLVNSSLHTPPAL 60
 DB 1 MASFTATAAAGSRWLGNGHNTOPPLSSOSSDLSYCSSLPMASRVTRKLVNSSLHTPPAL 60
 QY 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 DB 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 QY 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTHGFFDGDGMVHAV 180
 DB 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTHGFFDGDGMVHAV 180
 QY 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPKAIGELHGHGTGIARLMLFYARAAGIVDP 240
 DB 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPKAIGELHGHGTGIARLMLFYARAAGIVDP 240
 QY 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTVGRFDFDGLQESTMTIAHPKV 300
 DB 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTVGRFDFDGLQESTMTIAHPKV 300
 QY 301 DPESGELFALSYDVVSKPKYKFRSPDGTSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 DB 301 DPESGELFALSYDVVSKPKYKFRSPDGTSPDGTSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 QY 361 QVVFKLPEMIRGSPVVDKKNVAREGILDKYAEDSSNLIKWDADPCFCFHLNAAWEPE 420

```

|||||
Db 361 QVVKLPKPEMLRGSPVYDANKVARGFILLDKYAEDSSNIKWIDAPDFCFCHLNWAEPE 420
|||||
Qy 421 TDEVVVIGSCMTTPDSTIFNESDENLKSVLSEIRLNLTGSTRPIITSNEDQQVNLBAGM 480
|||||
Db 421 TDEVVVIGSCMTTPDSTIFNESDENLKSVLSEIRLNLTGSTRPIITSNEDQQVNLBAGM 480
|||||
Qy 481 VNRNMLGRKTKFAYLALAEWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPFLPGEGGE 540
|||||
Db 481 VNRNMLGRKTKFAYLALAEWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPFLPGEGGE 540
|||||
Qy 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPYFGHGTIFGADDLAKQVY 599
|||||
Db 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPYFGHGTIFGADDLAKQVY 599
|||||

RESULT 3
AAE04789
ID AAE04789 standard; Protein; 605 AA.
XX
AC AAE04789;
XX
DT 10-SEP-2001 (first entry)
XX
Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCE1.
XX
Tomato; neoxanthin cleavage enzyme; LeNCE1; abscisic acid; ABA;
XX stress tolerance; transgenic plant; plant breeding; antisense-therapy;
XX plant growth protectant; herbicide.
XX
OS Lycopersicon esculentum.
XX
PN EPL116794-A2.
XX
PD 18-JUL-2001.
XX
XX 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
XX
PR 11-JAN-2001; 2001JP-0003476.
XX
XX
XX (RIKE ) RIKEN KK.
XX
XX Iuchi S, Kobayashi M, Shinozaki K.
XX
XX WPI: 2001-400081/43.
XX
XX N-PSDB; AAD09401.
XX
XX
XX A DNA encoding a protein with a neoxanthin cleavage activity for
XX producing transgenic plants with improved or decreased stress tolerance
XX
XX Claim 3; Fig 2; 101pp; English.
XX
XX The invention relates to neoxanthin cleavage enzymes and their
XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
XX plant when expressed in a plant cell. The invention also relates to
XX methods for increasing or decreasing stress tolerance in a plant by
XX introducing the DNA into the plant, and a transgenic plant into which a
XX neoxanthin cleavage enzyme is introduced. The improvement of stress
XX tolerance in plants is useful, for example in plant breeding. Neoxanthin
XX cleavage enzyme genes are useful for producing transgenic plants. An arid
XX land can be improved by growing transformant weed for several years and
XX then removing the weed by specifically lowering stress tolerance in the
XX weed by inducing an inducible promoter. The present sequence is
XX Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCE1 protein
XX related to the invention.
XX
XX Sequence 605 AA;
XX
XX Query Match 72.4%; Score 2280.5; DB 22; Length 605;

```

```

Best Local Similarity 71.7%; Pred. No. 5 3e-222;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

Qy 5 TATAAAGSCRMGLGGNHTQPIPLSSOSSDLSYCS---SLPMASRVTRKLNVSALHTPTPALH 61
||| : || : || : || : || : || : || : || : || : || : || : || : ||
Db 3 TTTSHATNTWT-----KTKLSMPSSKEFFGAFASISLILKNQHRQSLNINSLSQAPLIIH 57
|||||
Qy 62 FPKOSSN--SPA--IVVKPKAKESN-----TKOMNLQRAAAALDAAEGFLVSHKLIHP 112
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 58 FPKOSSNYQTPKNTTISHPKOENNNSSSSTSKWNLVQKAAAMALDAVESALTTHEUEHP 117
|||||
Qy 113 LPTADPSVQIAGNAPVNEQPVRRNLPGVGLKPDLSIKGVYVRNGANPLHPBPVTHGHFFD 172
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 118 LPTADPRVQISGNAPVNPVPCOSLPVTKIPKCVQGVYVRNGANPLFEPTAGHFFD 177
|||||
Qy 173 GDGMVHAVKFEHGSASYACRFTQTNRFVQERQLGRPVFPKAIKELHGHGTIARLMLFYAR 232
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 178 GDGMVHAVQFKNGSASYACRFTETERLVQEKALGRPVFPKAIKELHGHGTIARLMLFYAR 237
|||||
Qy 233 AAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNGDLTKTVGRDFDGOLES 292
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 238 GLGLVDHSGKTGVANAGLVYFNNRLAMSEDDLPYHVKTPTGDLKTEGDFDGOLES 297
|||||
Qy 293 TMAHPKVDPSGELFALSVDWSKPYLYKFRFSPDGTSPDVEIQDQPTMMHDFATTE 352
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 298 TMAHPKLDPVSGELFALSVDYIQKPYLYKFRFSGKNGEKSNDELPVEDPTMMHDFATTE 357
|||||
Qy 353 NFVVVPOQVVKLPPEMIRGSPVYDKNKVARFGILDKYAEDSSNIKWIDAPDFCFHL 412
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 358 NFVVVPOQVVKPMSEIRGSPVYDKNKVSREGLDKYAKDGLKWEVDPDFCFHL 417
|||||
Qy 413 WNAWEEETDEVVVIGSCMTTPDSTIFNESDENLKSVLSEIRLNLTGSTRPIISNEDQ 472
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 418 WNAWEEAETDEIVVIGSCMTTPDSTIFNECDGLKSVLSEIRLNLTGSTRKSIENDE 477
|||||
Qy 473 QVNLEAGMVNENMLCRKTKFAYLALAEWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEP 532
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 478 QVNLEAGMVNENMLGRKTEYAYLALAEWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEP 537
|||||
Qy 533 FLPGEE--GGEDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPYFGHGTIFG 590
||||| : || : || : || : || : || : || : || : || : || : || : ||
Db 538 FLPRDPSKEEDDGYILAFVHDEREKWSELQIVNAVSLKLEATVKLPSRPYFGHGTIF 597
|||||
Qy 591 ADDLAKQ 597
|||||
Db 598 ANDLANQ 604
|||||

RESULT 4
AAB72308
ID AAB72308 standard; Protein; 605 AA.
XX
AC AAB72308;
XX
XX 16-MAY-2001 (first entry)
XX
XX Neoxanthin cleavage enzyme-like protein amino acid sequence.
XX
XX Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
XX NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
XX pathogen resistance; abscisic acid metabolism.
XX
XX Lycopersicon esculentum.
XX
XX WO200112801-A2.
XX
XX 22-FEB-2001.
XX
XX 17-AUG-2000; 2000WO-US22961.
XX
XX 18-AUG-1999; 99US-0149656.
XX
XX 23-MAY-2000; 2000US-0206405.
XX

```



```

Db 3 SSASNTWENALPSPFDFLSTSPNLLPL--RKTSSNTIICSLQI---LHPKQY 56
Qy 66 -----SSNSPAIVVKKAKESNT-----KOMNLFQRAAAALDAA 100
Db 57 OPTSTSTATTTPPIKTITTTTTPPRETNPLSDTNQPLPKQWFLQKAAATADLV 116
Qy 101 EGFVLSHEKHLPLKPTADPSVQIAGNEAPVNEOPVRRNLPVVGKLPDSTKGYVVRNGANP 160
Db 117 ETALYSHERKHLPLKPTADPSVQIAGNEAPVNEOPVRRNLPVVGKLPDSTKGYVVRNGANP 176
Qy 161 LHEPVTHHFFDGDGMVHAFKPEHGSASYACRFTOTNRFVQSRQLGRPVFPKAIAGELHGH 220
Db 177 LYEPVAGHHFFDGDGMVHAFKPEHGSASYACRFTOTNRFVQSRQLGRPVFPKAIAGELHGH 236
Qy 221 TGIARLMLFYARAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDLPYQVQITPNGDLKT 280
Db 237 SGIARLMLFYARGLFGLVDSQGMGVANAGLVYFNGRLLAMSEDDLPYHVRITPNGDLKT 296
Qy 281 VGRFDFDQLESTMLAHKPEKVPDSEGEFALSVDVYSKPYLYKVFRESPDGTSKSPDVEIQLD 340
Db 297 VGRYDFNGQLNSTMLAHKPEKVPDSEGEFALSVDVYSKPYLYKVFRESPDGTSKSPDVEIQLD 356
Qy 341 OPTMMHDAITENFVVDPQVQVFKLPEMIRGSPVVDKNNKVARFGLDKYAEADSSNIK 400
Db 357 EPTMMHDAITENFVVDPQVQVFKLPEMIRGSPVVDKNNKVARFGLDKYAEADSSNIK 416
Qy 401 WIDAPDCFCFHLNNAWEPETEVEVVIGSCMTPADSIENECESLKSVLSEIRLNLTKGE 460
Db 417 WIDAPDCFCFHLNNAWEPETEVEVVIGSCMTPADSIENECESLKSVLSEIRLNLTKGE 476
Qy 461 STRPILISNEQVNLAEAGMVRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLITIGEVKKH 520
Db 477 STRPILIS-DAFOVNLAEAGMVRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLITIGEVKKY 535
Qy 521 LYGNRYGGBLFLPCGEGEDEGYILCPVHDEKTKWSELQTVNVAUSLEVEATVKLPSRV 580
Db 536 MYGEKFGGEPFLP-NGOKEDDGYILAFVHDEKTKWSELQTVNVAUSLEVEATVKLPSRV 594
Qy 581 PYGRHGTFIGADDLAKQ 597
Db 595 PYGRHGTFIGADDLAKQ 611

RESULT 6
ABB91182
ID ABB91182 standard; Protein; 589 AA.
AC ABB91182;
XX ABB91182;
DT 31-MAY-2002 (first entry)
XX Herbicidally active polypeptide SEQ ID NO 393.
DE Herbicidally active polypeptide; herbicide.
KW Herbicidal; plant; agriculture; herbicide.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN WO200210210-A2.
PD 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP09892.
XX 28-AUG-2001; 2001WO-EP09892.
PA (FARB ) BAYER AG.
XX Tietjen K, Weidler M;
XX WPI; 2002-269010/31.
XX Identifying plant target proteins for herbicidally active compounds,

```

```

PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX Claim 5; SEQ ID NO 393; 261pp + Sequence Listing; English.
XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX Sequence 589 AA;
SQ
Query Match 64.9%; Score 2044.5; DB 23; Length 589;
Best Local Similarity 68.1%; Pred No. 4.7e-198;
Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps 4;
Qy 27 SSSSLVSCSSLPDMSRVIRKLVNSALHTPPEALHPFKOSSNSPAIVVKKAKESNTKOM 86
Db 27 SSSSVSFNTKPR---RRKLSANSVSDTPNLLNFPNYPSPNPII-----PEKDTSEW 76
Qy 87 NLFQRAAALDAAEGFLVSHKELHPLKPTADPSVQIAGNEAPVNEOPVRRNLPVVGKLP 146
Db 77 NLFQRAAALDAAEGFLVSHKELHPLKPTADPSVQIAGNEAPVNEOPVRRNLPVVGKLP 136
Qy 147 DSIKGVYVRNGANPLHEPVTHHFFDGDGMVHAFKPEHGSASYACRFTOTNRFVQSRQLG 206
Db 137 DCIDGVYVRNGANPLHEPVTHHFFDGDGMVHAFKPEHGSASYACRFTOTNRFVQSRQLG 196
Qy 207 RVFPKAIAGELHGHGFIARLMLFYARAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDL 266
Db 197 SPIFFKAIAGELHGHGFIARLMLFYARAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDL 256
Qy 267 PYQVQITPNGDLTKVGRFDFDQLESTMLAHKPEKVPDSEGEFALSVDVYSKPYLYKVFRES 326
Db 257 PYQVQITPNGDLTKVGRFDFDQLESTMLAHKPEKVPDSEGEFALSVDVYSKPYLYKVFRES 316
Qy 327 PDGTSKSPDVEIQLDQPTMMHDAITENFVVDPQVQVFKLPEMIRGSPVVDKNNKVARF 386
Db 317 PEGEKFGGEPFLP-NGOKEDDGYILAFVHDEKTKWSELQTVNVAUSLEVEATVKLPSRV 376
Qy 387 GILDKYAEADSSNIKALDAPDCFCFHLNNAWEPETEVEVVIGSCMTPADSIENECESLKS 446
Db 377 GILPRNAKDAEMVWVWVSPETFCFHLNNAWEPETEVEVVIGSCMTPADSIENECESLKS 436
Qy 447 SVLSEIRLNLTKGESTRPIISNEQVNLAEAGMVRNMLGRKTKFAYLALAEPPWPKVSG 506
Db 437 SVLSEIRLNLTKGESTRPIISNEQVNLAEAGMVRNMLGRKTKFAYLALAEPPWPKVSG 495
Qy 507 FAKVDLITIGEVKKHLYGNRYGGBLFLPCGEGEDEGYILCPVHDEKTKWSELQTVN 564
Db 496 FAKVDLITIGEVKKHLYGNRYGGBLFLPCGEGEDEGYILCPVHDEKTKWSELQTVN 555
Qy 565 AVSLEVEATVKLPSRVYFGHGTFIGADDLAKQ 597
Db 556 AVTLEATVKLPSRVYFGHGTFIGADDLAKQ 588

RESULT 7
AAE04782
ID AAE04782 standard; Protein; 583 AA.
XX AAE04782;
AC AAE04782;
XX 10-SEP-2001 (first entry)
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE1.
XX

```

KW Neoxanthin cleavage enzyme; AtNCE1; abscisic acid; ABA; herbicide;
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 KW plant growth protectant.
 XX Arabidopsis thaliana.
 XX EP1116794-A2.
 XX 18-JUL-2001.
 XX 11-JAN-2001; 2001EP-0300218.
 XX 13-JAN-2000; 2000JP-0010056.
 XX 11-JAN-2001; 2001JP-0003476.
 XX (RIKE) RIKEN KK.
 XX Iuchi S, Kobayashi M, Shinozaki K;
 DR WPI; 2001-400081/43.
 DR N-PSDB; AAD09394.
 XX A DNA encoding a protein with a neoxanthin cleavage activity for
 XX producing transgenic plants with improved or decreased stress tolerance
 XX Claim 3; Fig 10; 101pp; English.
 XX The invention relates to neoxanthin cleavage enzymes and their
 XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 XX plant when expressed in a plant cell. The invention also relates to
 XX methods for increasing or decreasing stress tolerance in a plant by
 XX introducing the DNA into the plant, and a transgenic plant into which a
 XX neoxanthin cleavage enzyme is introduced. The improvement of stress
 XX tolerance in plants is useful, for example in plant breeding. Neoxanthin
 XX cleavage enzyme genes are useful for producing transgenic plants. An arid
 XX land can be improved by growing transformant weed for several years and
 XX then removing the weed by specifically lowering stress tolerance in the
 XX weed by inducing an inducible promoter. The present sequence is
 XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE1 protein.
 XX The AtNCE1 cDNA is obtained from an Arabidopsis plant-derived cDNA
 XX library using a cDNA of the CPR65 (CowePea Responsive to Dehydration)
 XX gene isolated from cowpea plant as a probe.
 XX Sequence 583 AA;
 Query Match 63.2%; Score 1991; DB 22; Length 583;
 Best Local Similarity 64.5%; Pred. No. 1.2e-192;
 Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;
 QY 16 GGNHTQPLSSSGSDLSY--CSSLPMSRVTRKLNYSALHTPPALHFPKQSSNSPAIV 73
 DB 11 GGIKTWP-----QAQIDIGFRPIKQRPVKTQVIDVTE-LTKKRQLFTPTTATP--- 62
 QY 74 VKPKAKESNTKQNLFORAAAALDAEGFLVSHKLEHLPLKTDAPSVQIAGNFAPVNEQ 133
 DB 63 -----QHNFLRLNIFOKAARAIDAERALISHQSDPLPKTDAPRVQIAGNFSVPDES 116
 QY 134 PVRNRLPVGKLPDSIKGVYVRNGANPLHEFTVGHFFDGDGMVHAVKFEHGSASYACRF 193
 DB 117 SVRNLVTEGTIPDCIDGVYRNGANPMFEETAGHLLFDGDMVHAVKIINGSASYACRF 176
 QY 194 TQTNRFQERQLGRPVFPKAIIGELHGTIARLMLFYARAAGAVDPHAHGTGVANAGLVY 253
 DB 177 TKTERLVQEKRLGRPVFPKAIIGELHGSIGIARLMLFYVARGLCGLINQNGVGVANAGLVY 236
 QY 254 FNGELIAMSDDLPYQVQITPNDGLTKTVGRFDGQLESTMIAPKVPDSGSELFALSYD 313
 DB 237 FNNLLAMSDDLPYQVQITPNDGLTKTVGRFDGQLESTMIAPKVPDSGSELFALSYD 298
 QY 314 WSKPYLKYFRSPDGKSPDVEIQLDQPTMMHDAITENFVVVVDQOVVKLGEMISGK 373

DB 297 VVKPYLKYFRSPDGKSPDVEIQLDQPTMMHDAITENFVVVVDQOVVKLGEMISGK 356
 QY 374 SPVYDYKKNKVARFGILLDKYAEADSSNKKWIDAPDFCFHLNNAWBPETDEVVVGSCWTP 433
 DB 357 SPVYDGEKYSRLGIMPKDATEASQIIWVNSPETFCFHLNNAWBPETEEIVVIGSCMSP 416
 QY 434 PDSIFNESDENLKSVLSEIRLNLTGSTRRTPIISNEDQOVNLEAGMVRNMLGRKTKFA 493
 DB 417 ADSIFNERDESLSVLSIRINLRKTRKRLSLVNE--VNLEIGMVRNRLGRKTKFA 474
 QY 494 YLALAEPMKVSFGAKVDLTITTEVKKHLYGDNRYGGEPLFLPGEGG-----BEDGYYILCF 549
 DB 475 FLALAYPMKVSFGAKVDLCITGEMKKYIYGKYGGEFFLPGNSGNEEDDDGYIFCH 534
 QY 550 VHDEKTKWSELOIYNVNSLEVEATVKLPSRVPYGFHGTFFIGADDLAKQV 598
 DB 535 VHDEETKTSLOIINAVNLKLEATIKLPSRVPYGFHGTFFVDSNELVDQL 583
 RESULT 8
 AAE04788
 ID: AAE04788 standard; Protein; 504 AA.
 XX AAE04788;
 XX 10-SEP-2001 (first entry)
 XX Zea mays neoxanthin cleavage enzyme, VP14.
 DE Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 KW plant growth protectant.
 XX Zea mays.
 XX EP1116794-A2.
 XX 18-JUL-2001.
 XX 11-JAN-2001; 2001EP-0300218.
 XX 13-JAN-2000; 2000JP-0010056.
 XX 11-JAN-2001; 2001JP-0003476.
 XX (RIKE) RIKEN KK.
 XX Iuchi S, Kobayashi M, Shinozaki K;
 XX WPI; 2001-400081/43.
 XX N-PSDB; AAD09400.
 XX A DNA encoding a protein with a neoxanthin cleavage activity for
 XX producing transgenic plants with improved or decreased stress tolerance
 XX Claim 3; Fig 2; 101pp; English.
 XX The invention relates to neoxanthin cleavage enzymes and their
 XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 XX plant when expressed in a plant cell. The invention also relates to
 XX methods for increasing or decreasing stress tolerance in a plant by
 XX introducing the DNA into the plant, and a transgenic plant into which a
 XX neoxanthin cleavage enzyme is introduced. The improvement of stress
 XX tolerance in plants is useful, for example in plant breeding. Neoxanthin
 XX cleavage enzyme genes are useful for producing transgenic plants. An arid
 XX land can be improved by growing transformant weed for several years and
 XX then removing the weed by specifically lowering stress tolerance in the
 XX weed by inducing an inducible promoter. The present sequence is Zea mays
 XX neoxanthin cleavage enzyme, VP14 protein related to the invention.

SQ Sequence 604 AA;

Query Match 61.3%; Score 1930; DB 22; Length 604;
 Best Local Similarity 63.3%; Pred. No. 2.1e-186;
 Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

QY 22 PPLSSQSSDLSYCSLPMASRVTRKLNVSALHTP---PALHFPKQSSNSPAIVV--- 74
 DB 17 PARSARAS-----NSVRFSPRAVSSVPPAECLQAFHPKVADLPAPSKPAALAVPGHA 71

QY 75 -KPAKESNTKOMNLFQRAAAALDA-AEGFLVS-HEKHLPLPKTADPSVOIAGNFAPVN 131
 DB 72 AAPRAEGGKQLNLFQRAAAALDAFEEGVANVLERPHGLPSTADPAVOIAGNFAPVG 131

QY 132 EQPVRRNLPPVVGKLPDSLKGVYVRNGANPLHEPTVGTGHEFFDGDGMVHAVKEHGS-A 190
 DB 132 ERPPVHELPPVSGRIPPFDGVYARNGANPCDPVAGHLFDGDGMVHALRTRNGAESYA 191

QY 191 CRFTQTNRFVQEROLGRVPFPAKIGELHGHGTGIARLMLFYARAAGIYDPAHGTVANAG 250
 DB 192 CRFTETARLROERATGRVPFPAKIGELHGHGTGIARLMLFYARAAGIYDPAHGTVANAG 251

QY 251 LVYFNGRLAMSEDDLPYQVQITPNGDLKITVGRDFDQLESTMIHAKPKVDPSGELFAL 310
 DB 252 LVYFNGRLAMSEDDLPYHVRVADDDGLETGVGRYDFDQLESTMIHAKPKVDPSGELFAL 311

QY 311 SYDVWSKPYLKYFRSPDGTSPDVEIQLDQPTMHDFAITENFVVVPDQVQVFKLPEMI 370
 DB 312 SYDVIKRPLYKYFRSPDGTSPDVEIQLDQPTMHDFAITENFVVVPDQVQVFKLPEMI 371

QY 371 RGSFVYVYDKNKVARFGILDKYAEEDSNIKWIDAPDCFCFHLNNAWEEPTDEVVITGSC 430
 DB 372 RGSFVYVYDKNKVARFGILDKYAEEDSNIKWIDAPDCFCFHLNNAWEEPTDEVVITGSC 431

QY 431 MTPDPSIFNESDENLKSIVSEIRLNLTGSTRPIISNEDQOVNLEAGVMNRNMLGRKT 490
 DB 432 MTPDPSIFNESDENLKSIVSEIRLNLTGSTRPIISNEDQOVNLEAGVMNRNMLGRKT 490

QY 491 KFAYLALAEWPVKVSGFAKVDLTITGEVKKHLYGDNRYGGEPFLFPGRGGE-----EDEGY 545
 DB 546 ILCPVHDEKTKWSELOIVNAVSLVEATVKLPSRVPGFHGTFTICADDLAKQ 597
 DB 551 VLTEVHDEKTKWSELOIVNAVSLVEATVKLPSRVPGFHGTFTICADDLAKQ 602

RESULT 9
 ID AAB72309
 AC AAB72309 standard; Protein; 504 AA.
 AC AAB72309;
 XX 16-MAY-2001 (first entry)
 XX Neoxanthin cleavage enzyme-like protein amino acid sequence.
 XX Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
 KW NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
 KW pathogen resistance; abscisic acid metabolism.
 XX Zea mays.
 XX WO200112801-A2.
 XX 22-FEB-2001.
 XX 17-AUG-2000; 2000WO-US22961.
 XX 18-AUG-1999; 9905-0149656.
 XX 23-MAY-2000; 2000US-0206405.
 XX (PION-) PIONEER HT-BRED INT INC.

PA (CURA-) CURAGEN CORP.
 PT Bidney DL, Crasta OR, Hu X, Lu G;
 XX WPI; 2001-211215/21.
 XX Novel isolated defence-related signalling gene isolated from sunflower
 encoding neoxanthin cleavage enzyme, amino acid permease or glutamic
 acid-rich protein useful for increasing resistance of plant to a
 pathogen
 XX Disclosure; Fig 1; 135pp; English.
 XX This invention relates to defence-related signalling genes isolated from
 the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage
 enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich
 protein (GRP). The signalling gene is useful for increasing the
 resistance of a plant to a pathogen such as fungus, virus, bacterium,
 nematode or insect (e.g. European corn borer), preferably
Sclerotinia spp., *Phoma spp.*, or *Phomopsis spp.* by stably incorporating a
 construct containing the gene into the genome of the plant. The gene is
 useful for regulating gene expression in a plant, in response to a
 stimulus such as infection with a pathogen, damage from a pathogen,
 hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
 oxalic acid or expression of a gene encoding oxalic acid oxidase. The
 genes are also useful for stem-preferred regulation of gene expression in
 a plant. The genes are useful in agriculture, particularly in the
 breeding of crop plants with improved agronomic traits, for modifying
 abscisic acid (ABA) metabolism and for modifying amino acid transport and
 content in plants. The present sequence represents a neoxanthin cleavage
 enzyme-like protein from *Zea mays* used in the characterisation of
 sunflower NCE.
 XX Sequence 604 AA;
 SQ Query Match 61.3%; Score 1930; DB 22; Length 604;
 Best Local Similarity 63.3%; Pred. No. 2.1e-186;
 Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

Db 491 RYALVAEPKPSGFAKEDLSGELTKFEYGEGRGEGPCFVPMPPAAAHPRGEDGY 550
 QY 546 ILCFVHDEKTKWSELOIVNAVSEVATVKLPSPVPYGFHGTETIGADDLAKQ 597
 Db 551 VLFVHDERAGTSELLVVAADIRLEATVQLPSVPVGFHGTFTIGQELZQA 602

RESULT 10
 AAE04786
 ID AAE04786 standard; Protein; 577 AA.
 AC AAE04786;
 DT 10-SEP-2001 (first entry)
 XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5.
 KW Neoxanthin cleavage enzyme; AtNCED5; abscisic acid; ABA; herbicide;
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 KW plant growth protectant.
 OS Arabidopsis thaliana.
 XX
 XX EP1116794-A2.
 XX 18-JUL-2001.
 XX 11-JAN-2001; 2001EP-0300218.
 XX 13-JAN-2000; 2000JP-0010056.
 PR 11-JAN-2001; 2001JP-0003476.
 XX (RIKE) RIKEN KK.
 PI Iuchi S, Kobayashi M, Shinozaki K;
 XX WPI; 2001-400081/43.
 DR N-PSDB; AAD09398.
 PT A DNA encoding a protein with a neoxanthin cleavage activity for
 PT producing transgenic plants with improved or decreased stress tolerance
 PT
 XX Claim 3; Fig 10; 101pp; English.
 CC The invention relates to neoxanthin cleavage enzymes and their
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 CC plant when expressed in a plant cell. The invention also relates to
 CC methods for increasing or decreasing stress tolerance in a plant by
 CC introducing the DNA into the plant, and a transgenic plant in which a
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid
 CC land can be improved by growing transgenic plants. An arid
 CC then removing the weed by specifically lowering stress tolerance in the
 CC weed by inducing an inducible promoter. The present sequence is
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein.
 CC The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA
 CC library using a cDNA of the CPRD65 (CowPea Responsive to Dehydration)
 CC gene isolated from cowpea plant as a probe.
 XX Sequence 577 AA;
 SQ

Query Match
 Best Local Similarity 52.8%; Score 1663.5; DB 22; Length 577;
 Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;
 QY 24 LSSQSSDLSCSSLPASRVTRKLNVS-ALHTPPALHFPKQSSNAIVVKPKAKESN 82
 Db 9 LLPTKTSRSHLLPOPKNANISRRILNPKFTPLDLSVPSP-----VKLKPYPN 62

QY 83 TKQNLFORAAAALDAAE-GFLVSHKHLHPKPTADPSVQIAGNFAPVNEQPVRRNLPV 141
 Db 63 ---LNLQKLAATMLDKIESIVIPMEQNRPKPPTDPAVLSCGNFAPVNECPVQNGLEV 119
 QY 142 VGKLPDSIKGVYVRNGANPLHEPVTHGHFDDGDMHAAVK--FEHGSASACRTOINRF 199
 Db 120 VGOIPSLKGVYIRNGANPMFPPLAGHHLFDGDMHAAVSGFDN-QVSYSCRYTKTNRL 178
 QY 200 VQERQLGRPVFPKAIAGELHGHGTIARLMLFYARAAAAGIVDPAHGTGVANAGLVVFNRL 259
 Db 179 VQETALGRSVFPPKPIGELHGHSLARLALFTARAGICLVDSGRMGVANAGVVFNGRL 238
 QY 260 AMSDDLPIYOQVITPNCGLKTVGRFDEGQLESTMIAPKVPDESSELFALSVDVSKPY 319
 Db 239 ANSEDDLPYQVKIDGGQDLETIGREFDDQIDSSVIAHPKVDATTDGLHTLSYNVLKKPH 298
 QY 320 LKYFRSPDGTGKSPDVEIQDQPTMMDFAITENFVVVDDQVQVFKLPEMIRGSSPVYD 379
 Db 299 LRYLKENTCGKTRDVEITLPEPTIMHDAITENFVVVDDQVQVFKLSEMRGSPVIYV 358
 QY 380 KKVAREGILDKYAESSNIKWDADPCFCFHLNAAWEEPETDE---VVVIGSCMTTPD 435
 Db 359 KEMARFVLUSKQDLTGSDINWYDVPDCFCFHLNAAWEE-RTEEGDPVIVVIGSCMSPD 417
 QY 436 SIFNESDENLKSVLSEIRLNKLTGSTRPPIISNEQQVNLNLEAGMVRNMLGRKTKFAYL 495
 Db 418 TIFSESEPTRVSEIRLNKMTKESNRKVIVTG---VNLEAGHNRSTVGRKSOVFYI 473
 QY 496 ALAEPKPVSGFAKVDLTGTGGEKKHLYGDNRYGGEPLFLFEGGEDEGYILCFVHDEKT 555
 Db 474 AIADPWKSGIAKVDIQNTGVSEFNYPGSRFGGECPCFVPEGGEEDKGYVMGCFVRDEK 533
 QY 556 WKSELOIVNAVSEVATVKLPSPVPYGFHGTETIGADDLAKQV 598
 Db 534 DESEFVVVDATDMKQAAVAKLPERPYPYGFHGTETVSENLKEQV 576

RESULT 11
 ABB92416
 ID ABB92416 standard; Protein; 577 AA.
 XX ABB92416;
 DT 31-MAY-2002 (first entry)
 DE Herbicidally active polypeptide SEQ ID NO 1627.
 KW Herbicidal; plant; agriculture; herbicide.
 XX Arabidopsis thaliana.
 XX WO200210210-A2.
 PD 07-FEB-2002.
 PF 28-AUG-2001; 2001WO-EP09892.
 XX 28-AUG-2001; 2001WO-EP09892.
 PA (FARB) BAYER AG.
 PI Tietjen K, Weidner M;
 XX WPI; 2002-269010/31.
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 XX organisms -
 PS Claim 5; SEQ ID NO 1627; 26lpp + Sequence Listing; English.
 XX The invention relates to identifying target proteins
 CC

CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX SQ Sequence 577 AA;

Query Match 52.8%; Score 1663.5; DB 23; Length 577;
 Best Local Similarity 55.4%; Pred. No. 2.2e-159;
 Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

QY 24 LSSQSSDLSYSSCPMASTRVKLNVS-ALHTPPALHFKQSSNPAIVVKKPAKESN 82
 Db 9 LLPTKISRSHLLPOPKVANSRRLLINPKTIPTLDTSPVSP-----VKLKPTYPN 62
 QY 83 TKQMLFORAAALDAAE-GFLVSHKHLPLKPTADPSVQIAGNFAVPNEQPVRENLPV 141
 Db 63 ---LNLQKLAATLCKTSSIVIPMEQNRPLPKPTDPAVQLSGNFAVNECPVONGLEV 119
 QY 142 VGKLPDSIKGVYVRNGANPLHEPVTGHHFTGDDGMVHAVK--FEHGSASYACRFQTNR 199
 Db 120 VQIPSCULKGVYIRNGANPMFPPLAGHHLFDGDMHIAVSGFDN-QVSYSCRYTKTKRL 178
 QY 200 VQERQLGRVPKPAIGELHGHGTGLARLMLFVAAGATVDPAHGTGVANAGLVFNGRL 259
 Db 179 VQETALGRSVFPKPIGELHGHGTGLARLALFTARAGIGLVDTGRMGVANAGVVFNGRL 238
 QY 260 AMSDDLPLQYQVITPNGDLKTVGRFDDGQLESTMIARHPKVDPRSGELFALSYDWSKPY 319
 Db 239 AMSDDLPLQYQKIDGQDLEITGRFPDQIDSSVIAHPKYDATIGDLHTLSYNVLKPKH 298
 QY 320 LKYEFPDGTGSDVELQDPTMMHDEATENFVVPDQOVVFKLPFEMIRGGSPVYVD 379
 Db 299 LRYLFTNGCKKTRDVEITLPEPTMIHDEATENFVVPDQOVVFKLSEMIIRGGSPVYV 358
 QY 380 KNKVAREGILDKYAEDSSNKKWIDAPDCFCFLHNAEEPTDE---VVVIGSCMPDP 435
 Db 359 KEKARGVLSKQDLTSDINWVDVPCFCFLHNAEE-RTGEGDPVIVVIGSCMPPD 417
 QY 436 SFNEDENLKSVLSEIRLNLKLGESTRRPIISNEDQVNLGAGVNRNMLGRKTKFAYL 495
 Db 418 TIFSEGEPTIRVELSEIRLNRKTESNKKVITVG---VNLGAGHINRSYGRKSQFYI 473
 QY 496 ALAEPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLPLPGEGGEDEGYILCFVHDEK 555
 Db 474 AIADPWPKSGIAKVDIQNTGVSEFNFGPGRFGGECFVPEGEEDKGYMGEVGRDEEK 533
 QY 556 WKSEIQIVNAVSLEAVEATVVKLPSPVYPYGFHGTFTIGADDLAKQV 598
 Db 534 DESEFVVVDATDMKQVAARLPERVPYGFHGTFTVSENLQEQV 576

RESULT 12

AAG31333
 ID AAG31333 standard; Protein; 595 AA.

XX AC AAG31333;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37612.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX SS EP1033405-A2.

PN

XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 03-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.

BEST AVAILABLE COPY

```
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.7%; Score 968; DB 21; Length 595;
Best Local Similarity 36.7%; Pred. No. le-89;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLVSSALHTPPALHEPKQSSNPAIVVKKAKESNTKOMNLFORAAAAADAAE 101
DB 42 SPITNPSSNDNRNPKTLH---NRTNHTLVSSPPKLRPEMTLALF---TTVEDVIN 94
QY 102 GFLVSHEKHLPLKPTADPSVQIAGNFAPVNEQPVRRNLPV--GKLPDSIKGVVRNGANP 160
DB 95 TEIDP-----PSRPSVDPKHVLSDNFAPVLDLPPTDCEIIHGTLPLSLNGAYIRGNPN 149
QY 161 LHEPVTGHEFFDGDGMVHAKVEHGSASVACRFTOTNRVQBRQLGRPVFPKAIHELGH 220
DB 150 QFLPRGPHLFDGDMGLHAIKHNGKATLCRSYVTKYKYNVEKQTGAPVMPNVFSGFNCV 209
QY 221 T-GIARLMFYARAAAGIVDPAHGTGVANAGLVYENGRLIAMEDDLPOVQOITNGDLK 279
DB 210 TASVARGALTAARVLITGQYNPNVNGIGLANTSIAFFSNRLFALGESDLPYAVALTESGDIE 269
QY 280 TVGREFDGOOLESTMIAPKVPDPSGELFALSVDVVKPYLKYFRFSPDGTSPDVEI-Q 338
DB 270 TIGRYDFDGLKAMSTAHPKTDPITGETFAFYGVV-PPFLTYFRFDSAGKKQORDVPIFS 328
QY 339 LDQPTMHDFALTENFVVVPDQVYFK---LPETIRGGSPVYVYDKNKNVAREFILDKYAED 395
DB 329 MTSFSLHDFAITKRHAIFAETQLGMRNMLDLVLEGGSPVCTDNGKTPRLGVIPKYAGD 388
QY 396 SSNIKWIDAPDCFCFHLNANEEPEPTDEVVVGSCMTTPDSIFNDESDENLKSVLSEIKLN 455
DB 389 ESEMAWFEVPGNIITHAINAWDEDDGNSVLAPNIMIEHTLERMD-LVHALVEKVKID 447
QY 456 LKTGESIRRPRIISNEDQOVNLEAGMVNRNMLGRKTKFAYLALAEWPVKVVSFAKVDLTG 515
```

```

Db 448 LVTGIVRRHPISAR-----NLDFAVINPAFLGRCGRVYVAALGDPMPKISGVVKLDVSGK 502
QY 516 E-----VKKHLYGDNRYGGEPLFPGEGG-----EDEGVILCFVDEKTKWSELQIVNAV 566
Db 503 DRDDCTVARMYIGSCYCGEPEFFVARDPONGPRAEEDDGYVTVYVHDEVGTGSKFLVMDAK 562
QY 567 S--LEVEATVKLPSRPVPGFHGTGACDLAK 596
Db 563 SPELEIVAVALPRVPYGFHGLFVKESDLNK 594

RESULT 13
AAE04783
ID AAE04783 standard; Protein; 595 AA.
XX AC
XX AC
XX AAE04783;
DT 10-SEP-2001 (first entry)
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE2.
XX Neoxanthin cleavage enzyme; AtNCE2; abscisic acid; ABA; herbicide;
XX stress tolerance; transgenic plant; plant breeding; antisense-therapy;
XX plant growth protectant.
XX Arabidopsis thaliana.
XX OS
XX EP1116794-A2.
XX 18-JUL-2001.
XX PF 11-JAN-2001; 2001EP-0300218.
XX PR 13-JAN-2000; 2000JP-0010056.
XX PR 11-JAN-2001; 2001JP-0003476.
XX PA (RIKE ) RIKEN KK.
XX PI Iuchi S, Kobayashi M, Shinozaki K;
XX N-PSDB; AAD09395.
XX A DNA encoding a protein with a neoxanthin cleavage activity for
XX producing transgenic plants with improved or decreased stress tolerance
XX
XX Example 10; Fig 10; 101pp; English.
XX The invention relates to neoxanthin cleavage enzymes and their
XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
XX plant when expressed in a plant cell. The invention also relates to
XX methods for increasing or decreasing stress tolerance in a plant by
XX introducing the DNA into the plant, and a transgenic plant into which a
XX neoxanthin cleavage enzyme is introduced. The improvement of stress
XX tolerance in plants is useful, for example in plant breeding. Neoxanthin
XX cleavage enzyme genes are useful for producing transgenic plants. An arid
XX land can be improved by growing transformant weed for several years and
XX then removing the weed by specifically lowering stress tolerance in the
XX Arabidopsis thaliana neoxanthin cleavage enzyme. The present sequence is
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE2 protein.
XX The AtNCE2 cDNA is obtained from an Arabidopsis plant-derived cDNA
XX library using a cDNA of the CPR65 (CowPea Responsive to dehydration)
XX gene isolated from cowpea plant as a probe.
XX Sequence 595 AA;

```

Query Match 30.7%; Score 968; DB 22; Length 595;
 Best Local Similarity 36.7%; Pred. No. 1e-88;
 Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

```

QY 42 SRVTRKLVSSALHTPPALHFFKQSSNPAIVVVKPAKESNTKOMLQFRAAAALDAAE 101
Db 42 SPITNPSONDRNRKPKTLH---NRTNHTLVSPPKLRPEMTLATALF-----TTVEDVIN 94
QY 102 GLVLSHEKHLPLPKTADPSVQIAGNFAPVNEQFRRNLVV--GKLPSDSIKGVYVRNGANP 160
Db 95 TFDIP-----PSRPSVDPKHVLSDNFAPVLDLPDCEIIGHITLPLSLNGAYIRNGPNP 149
QY 161 LHEPVTGHHFFDGDGMVHAVKFEHGSASACFTOTNRVQERQLGRVPFPAIGELHGH 220
Db 150 QFLPRGPYHLFDGDMHLHAIKHNGKATLCRSYVTVKYNVBKQTGAPVMPVNFSGNGV 209
QY 221 T-GIARLMLFYARAAGIVDPAHGTGVANAGLVYFNGRIAMSEDDLPYQVOITFNGDLK 279
Db 210 TASVARGALTAARVLGTQVNPVNGIGLANTSLAFTSNRLFALGESDLPYAVELTSGDIE 269
QY 280 TVGREDEDCOLESTMIAPKVPDPESGELFALSYDVVSKPYLYKFRFSPDGTKSPDVEI-Q 338
Db 270 TIGRYDEFDCKUAMSMIAHPKTDPTITGETFAFRYGPV-PPFLTYFRFDSAGKKQDVPIFS 328
QY 339 LDQPTMHDDFAITENFVVVDOOVVFK---LPEMIRGGSPVYVDKKNKYARFGILDKYAD 395
Db 329 MTPSPFLHDFAITKRAHFAETQLGWRNMMLDLVLEGGSPVGTGDKTGRUGVIPKYAGD 388
QY 396 SSNIKWDAPDCCFHLNNAMEPETDEVVIGVSCMTPDPSIFNESDENLKSVLSEIRLN 455
Db 389 EEMKWFEPVPGFNIIHAINAWDEDDGNSVLIAPNIMISIEHTLRMD-LVHALVEKVKID 447
QY 456 LKGTGSTRRPILSNEDQOVNLEAGVNVNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGT 515
Db 448 LVTGIVRRHPISAR-----NLDFAVINPAFLGRCGRVYVAALGDPMPKISGVVKLDVSGK 502
QY 516 E-----VKKHLYGDNRYGGEPLFPGEGG-----EDEGVILCFVDEKTKWSELQIVNAV 566
Db 503 DRDDCTVARMYIGSCYCGEPEFFVARDPONGPRAEEDDGYVTVYVHDEVGTGSKFLVMDAK 562
QY 567 S--LEVEATVKLPSRPVPGFHGTGACDLAK 596
Db 563 SPELEIVAVALPRVPYGFHGLFVKESDLNK 594

```

RESULT 14

AAE04783
 ID AAE04783 standard; Protein; 595 AA.
 XX AC
 XX AAE04783;
 DT 16-MAY-2001 (first entry)
 XX Arabidopsis thaliana.

Neoxanthin cleavage enzyme-like protein amino acid sequence.

Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
 NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
 pathogen resistance; abscisic acid metabolism.

Arabidopsis thaliana.

WC200112801-A2.

22-FEB-2001.

17-AUG-2000; 2000WO-0522961.

18-AUG-1999; 99US-0149656.

23-MAY-2000; 2000US-0206405.

(PION-) PIONEER HI-BRED INT INC.

(CURA-) CURAGEN CORP.

Bidney DL, Crasta OR, Hu X, Lu G;

WPI; 2001-211215/21.

PR	18-JUN-1999;	9905-0139457.
PR	18-JUN-1999;	9905-0139458.
PR	18-JUN-1999;	9905-0139459.
PR	18-JUN-1999;	9905-0139460.
PR	18-JUN-1999;	9905-0139461.
PR	18-JUN-1999;	9905-0139462.
PR	18-JUN-1999;	9905-0139463.
PR	18-JUN-1999;	9905-0139470.
PR	18-JUN-1999;	9905-0139475.
PR	21-JUN-1999;	9905-0139763.
PR	21-JUN-1999;	9905-0139817.
PR	23-JUN-1999;	9905-0139899.
PR	23-JUN-1999;	9905-0140353.
PR	23-JUN-1999;	9905-0140354.
PR	24-JUN-1999;	9905-0140695.
PR	28-JUN-1999;	9905-0140823.
PR	28-JUN-1999;	9905-0140991.
PR	30-JUN-1999;	9905-0141287.
PR	01-JUL-1999;	9905-0141842.
PR	01-JUL-1999;	9905-0142154.
PR	02-JUL-1999;	9905-0142055.
PR	02-JUL-1999;	9905-0142390.
PR	08-JUL-1999;	9905-0142803.
PR	09-JUL-1999;	9905-01442920.
PR	12-JUL-1999;	9905-01442977.
PR	13-JUL-1999;	9905-0144352.
PR	14-JUL-1999;	9905-01443624.
PR	15-JUL-1999;	9905-01444005.
PR	16-JUL-1999;	9905-01444085.
PR	16-JUL-1999;	9905-01444086.
PR	19-JUL-1999;	9905-01444325.
PR	19-JUL-1999;	9905-01444331.
PR	19-JUL-1999;	9905-01444332.
PR	19-JUL-1999;	9905-01444333.
PR	19-JUL-1999;	9905-01444334.
PR	20-JUL-1999;	9905-01444352.
PR	20-JUL-1999;	9905-01444352.
PR	20-JUL-1999;	9905-01446824.
PR	21-JUL-1999;	9905-0144814.
PR	21-JUL-1999;	9905-0145086.
PR	21-JUL-1999;	9905-0145088.
PR	22-JUL-1999;	9905-0145085.
PR	22-JUL-1999;	9905-0145087.
PR	22-JUL-1999;	9905-0145089.
PR	22-JUL-1999;	9905-0145192.
PR	23-JUL-1999;	9905-0145145.
PR	23-JUL-1999;	9905-0145218.
PR	23-JUL-1999;	9905-0145224.
PR	26-JUL-1999;	9905-0145276.
PR	27-JUL-1999;	9905-0145913.
PR	27-JUL-1999;	9905-0145918.
PR	27-JUL-1999;	9905-0145919.
PR	28-JUL-1999;	9905-0145951.
PR	02-AUG-1999;	9905-0146386.
PR	02-AUG-1999;	9905-0146388.
PR	02-AUG-1999;	9905-0146389.
PR	03-AUG-1999;	9905-0147208.
PR	04-AUG-1999;	9905-0147302.
PR	04-AUG-1999;	9905-0147302.
PR	05-AUG-1999;	9905-0147352.
PR	05-AUG-1999;	9905-0147260.
PR	06-AUG-1999;	9905-0147303.
PR	06-AUG-1999;	9905-0147416.
PR	09-AUG-1999;	9905-0147419.
PR	09-AUG-1999;	9905-0147493.
PR	09-AUG-1999;	9905-0147935.
PR	10-AUG-1999;	9905-0148171.
PR	11-AUG-1999;	9905-0148319.
PR	12-AUG-1999;	9905-0148341.
PR	13-AUG-1999;	9905-0148565.
PR	13-AUG-1999;	9905-0148684.
PR	16-AUG-1999;	9905-0149368.
PR	17-AUG-1999;	9905-0149175.
PR	18-AUG-1999;	9905-0149426.

PR	20-AUG-1999	99US-01497221
PR	20-AUG-1999	99US-0149723
PR	20-AUG-1999	99US-0149729
PR	20-AUG-1999	99US-0149929
PR	23-AUG-1999	99US-0149902
PR	23-AUG-1999	99US-0149930
PR	25-AUG-1999	99US-0150566
PR	25-AUG-1999	99US-0150884
PR	27-AUG-1999	99US-0151065
PR	27-AUG-1999	99US-0151066
PR	27-AUG-1999	99US-0151080
PR	30-AUG-1999	99US-0151303
PR	31-AUG-1999	99US-0151438
PR	01-SEP-1999	99US-0151930
PR	07-SEP-1999	99US-0152363
PR	10-SEP-1999	99US-0153070
PR	13-SEP-1999	99US-0153758
PR	13-SEP-1999	99US-0154018
PR	16-SEP-1999	99US-0154039
PR	20-SEP-1999	99US-0154779
PR	22-SEP-1999	99US-0155139
PR	23-SEP-1999	99US-0155148
PR	24-SEP-1999	99US-0155659
PR	28-SEP-1999	99US-0156458
PR	28-SEP-1999	99US-0156596
PR	04-OCT-1999	99US-0157117
PR	04-OCT-1999	99US-0157753
PR	06-OCT-1999	99US-0158765
PR	07-OCT-1999	99US-0158029
PR	08-OCT-1999	99US-0158232
PR	12-OCT-1999	99US-0158369
PR	13-OCT-1999	99US-0159233
PR	13-OCT-1999	99US-0159294
PR	13-OCT-1999	99US-0159295
PR	14-OCT-1999	99US-0159329
PR	14-OCT-1999	99US-0159331
PR	14-OCT-1999	99US-0159637
PR	14-OCT-1999	99US-0159638
PR	18-OCT-1999	99US-0159584
PR	21-OCT-1999	99US-0160741
PR	21-OCT-1999	99US-0160767
PR	21-OCT-1999	99US-0160768
PR	21-OCT-1999	99US-0160770
PR	21-OCT-1999	99US-0160814
PR	22-OCT-1999	99US-0160815
PR	22-OCT-1999	99US-0160880
PR	22-OCT-1999	99US-0160981
PR	22-OCT-1999	99US-0160989
PR	25-OCT-1999	99US-0161404
PR	25-OCT-1999	99US-0161405
PR	25-OCT-1999	99US-0161406
PR	26-OCT-1999	99US-0161359
PR	26-OCT-1999	99US-0161360
PR	26-OCT-1999	99US-0161361
PR	28-OCT-1999	99US-0161920
PR	28-OCT-1999	99US-0161992
PR	28-OCT-1999	99US-0161993
PR	29-OCT-1999	99US-0162142

Query Match 30.48; Score 959; DB 21; Length 517;

Best Local Similarity 39.2%; Pred. No. 6.5e-88;

BEST LOCAL SIMILARITY	55.2%	PREC. NO.	0.56	007
Matches	197;	Conservative	98;	Mismatches 183;
			Indels	24; Gaps 10;

[illegible]

Db 142 AARVITGOYNPVNGIGLANTSLAFFSNLFAIGESDLPYAVRLTESGDIETIGRYDEGK 201
QY 290 LESTIAHPKVDPSGELFALSVDWVKPYLKYPRFSPDGTKSPDVEI-QLDQPTMMHDF 348
Db 202 LAMSTAHPKTDPITGETFAERYGPV-PPFLTYPRFDSAGKKQORDVPFSTSPSFLHDF 260
QY 349 AITENFVVVPDQOVVK---LPEMIRGGSPVYDKNKVAREGILDKYAEDSSNIKWIDAP 405
Db 261 AITKRHAIFAIOIGMRNMMLDLVLEGGSPVGTONGKTPRLGVTPKYAGDESEMKWFVP 320
QY 406 DCFCFHLNNAWEEPETDEVVVIGSCHMTPPDSIFNESDENLKSIVLSEIRLNLKTGESTRRP 465
Db 321 GFNIHAINAWEDDGSVLIAPNIMSIEHTLERMD-LVHALVEKVKIDLVITGIVRHP 379
QY 466 IISNEDQOVNLEAGMVRNMLGRKTKFAYLALAEFPWKVSGFAKVDLTJGE-----VKKH 520
Db 380 ISAR-----NLDEFAVINPAELGRCRSRVYAAIGDPMKISGVVKLDYSKGRDDDCIVARR 434
QY 521 LYGNRYGGEPLFLPGEGG----BEDGYILCFVHDEKTKWSEIQVNAVS--LEVEATV 574
Db 435 MYGCGYGGEPFFVARDPGNPEAEEDDGYVYVHDEVGTGESKFLVMDAKSPELEIVAAY 494
QY 575 KLPSRPYGFHGTFIGADDLAK 596
Db 495 RLPRRPYGFHGLFVKESDLNK 516

Search completed: May 19, 2003, 15:47:11
Job time : 45 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:15 ; Search time 29 Seconds
(without alignments)
607.736 Million cell updates/sec

Title: US-09-758-269-6
 Perfect score: 3150
 Sequence: 1 MASFTAAVAVSGRWLGGNHT.....VPYGFHGTFFIGADDLAKQV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Do not use the following symbols: π , ∞ , $\frac{1}{2}$, $\frac{3}{4}$, $\frac{5}{6}$, $\frac{7}{8}$, $\frac{9}{10}$, $\frac{11}{12}$, $\frac{13}{14}$, $\frac{15}{16}$, $\frac{17}{18}$, $\frac{19}{20}$, $\frac{21}{22}$, $\frac{23}{24}$, $\frac{25}{26}$, $\frac{27}{28}$, $\frac{29}{30}$, $\frac{31}{32}$, $\frac{33}{34}$, $\frac{35}{36}$, $\frac{37}{38}$, $\frac{39}{40}$, $\frac{41}{42}$, $\frac{43}{44}$, $\frac{45}{46}$, $\frac{47}{48}$, $\frac{49}{50}$, $\frac{51}{52}$, $\frac{53}{54}$, $\frac{55}{56}$, $\frac{57}{58}$, $\frac{59}{60}$, $\frac{61}{62}$, $\frac{63}{64}$, $\frac{65}{66}$, $\frac{67}{68}$, $\frac{69}{70}$, $\frac{71}{72}$, $\frac{73}{74}$, $\frac{75}{76}$, $\frac{77}{78}$, $\frac{79}{80}$, $\frac{81}{82}$, $\frac{83}{84}$, $\frac{85}{86}$, $\frac{87}{88}$, $\frac{89}{90}$, $\frac{91}{92}$, $\frac{93}{94}$, $\frac{95}{96}$, $\frac{97}{98}$, $\frac{99}{100}$, $\frac{101}{102}$, $\frac{103}{104}$, $\frac{105}{106}$, $\frac{107}{108}$, $\frac{109}{110}$, $\frac{111}{112}$, $\frac{113}{114}$, $\frac{115}{116}$, $\frac{117}{118}$, $\frac{119}{120}$, $\frac{121}{122}$, $\frac{123}{124}$, $\frac{125}{126}$, $\frac{127}{128}$, $\frac{129}{130}$, $\frac{131}{132}$, $\frac{133}{134}$, $\frac{135}{136}$, $\frac{137}{138}$, $\frac{139}{140}$, $\frac{141}{142}$, $\frac{143}{144}$, $\frac{145}{146}$, $\frac{147}{148}$, $\frac{149}{150}$, $\frac{151}{152}$, $\frac{153}{154}$, $\frac{155}{156}$, $\frac{157}{158}$, $\frac{159}{160}$, $\frac{161}{162}$, $\frac{163}{164}$, $\frac{165}{166}$, $\frac{167}{168}$, $\frac{169}{170}$, $\frac{171}{172}$, $\frac{173}{174}$, $\frac{175}{176}$, $\frac{177}{178}$, $\frac{179}{180}$, $\frac{181}{182}$, $\frac{183}{184}$, $\frac{185}{186}$, $\frac{187}{188}$, $\frac{189}{190}$, $\frac{191}{192}$, $\frac{193}{194}$, $\frac{195}{196}$, $\frac{197}{198}$, $\frac{199}{200}$, $\frac{201}{202}$, $\frac{203}{204}$, $\frac{205}{206}$, $\frac{207}{208}$, $\frac{209}{210}$, $\frac{211}{212}$, $\frac{213}{214}$, $\frac{215}{216}$, $\frac{217}{218}$, $\frac{219}{220}$, $\frac{221}{222}$, $\frac{223}{224}$, $\frac{225}{226}$, $\frac{227}{228}$, $\frac{229}{230}$, $\frac{231}{232}$, $\frac{233}{234}$, $\frac{235}{236}$, $\frac{237}{238}$, $\frac{239}{240}$, $\frac{241}{242}$, $\frac{243}{244}$, $\frac{245}{246}$, $\frac{247}{248}$, $\frac{249}{250}$, $\frac{251}{252}$, $\frac{253}{254}$, $\frac{255}{256}$, $\frac{257}{258}$, $\frac{259}{260}$, $\frac{261}{262}$, $\frac{263}{264}$, $\frac{265}{266}$, $\frac{267}{268}$, $\frac{269}{270}$, $\frac{271}{272}$, $\frac{273}{274}$, $\frac{275}{276}$, $\frac{277}{278}$, $\frac{279}{280}$, $\frac{281}{282}$, $\frac{283}{284}$, $\frac{285}{286}$, $\frac{287}{288}$, $\frac{289}{290}$, $\frac{291}{292}$, $\frac{293}{294}$, $\frac{295}{296}$, $\frac{297}{298}$, $\frac{299}{300}$, $\frac{301}{302}$, $\frac{303}{304}$, $\frac{305}{306}$, $\frac{307}{308}$, $\frac{309}{310}$, $\frac{311}{312}$, $\frac{313}{314}$, $\frac{315}{316}$, $\frac{317}{318}$, $\frac{319}{320}$, $\frac{321}{322}$, $\frac{323}{324}$, $\frac{325}{326}$, $\frac{327}{328}$, $\frac{329}{330}$, $\frac{331}{332}$, $\frac{333}{334}$, $\frac{335}{336}$, $\frac{337}{338}$, $\frac{339}{340}$, $\frac{341}{342}$, $\frac{343}{344}$, $\frac{345}{346}$, $\frac{347}{348}$, $\frac{349}{350}$, $\frac{351}{352}$, $\frac{353}{354}$, $\frac{355}{356}$, $\frac{357}{358}$, $\frac{359}{360}$, $\frac{361}{362}$, $\frac{363}{364}$, $\frac{365}{366}$, $\frac{367}{368}$, $\frac{369}{370}$, $\frac{371}{372}$, $\frac{373}{374}$, $\frac{375}{376}$, $\frac{377}{378}$, $\frac{379}{380}$, $\frac{381}{382}$, $\frac{383}{384}$, $\frac{385}{386}$, $\frac{387}{388}$, $\frac{389}{390}$, $\frac{391}{392}$, $\frac{393}{394}$, $\frac{395}{396}$, $\frac{397}{398}$, $\frac{399}{400}$, $\frac{401}{402}$, $\frac{403}{404}$, $\frac{405}{406}$, $\frac{407}{408}$, $\frac{409}{410}$, $\frac{411}{412}$, $\frac{413}{414}$, $\frac{415}{416}$, $\frac{417}{418}$, $\frac{419}{420}$, $\frac{421}{422}$, $\frac{423}{424}$, $\frac{425}{426}$, $\frac{427}{428}$, $\frac{429}{430}$, $\frac{431}{432}$, $\frac{433}{434}$, $\frac{435}{436}$, $\frac{437}{438}$, $\frac{439}{440}$, $\frac{441}{442}$, $\frac{443}{444}$, $\frac{445}{446}$, $\frac{447}{448}$, $\frac{449}{450}$, $\frac{451}{452}$, $\frac{453}{454}$, $\frac{455}{456}$, $\frac{457}{458}$, $\frac{459}{460}$, $\frac{461}{462}$, $\frac{463}{464}$, $\frac{465}{466}$, $\frac{467}{468}$, $\frac{469}{470}$, $\frac{471}{472}$, $\frac{473}{474}$, $\frac{475}{476}$, $\frac{477}{478}$, $\frac{479}{480}$, $\frac{481}{482}$, $\frac{483}{484}$, $\frac{485}{486}$, $\frac{487}{488}$, $\frac{489}{490}$, $\frac{491}{492}$, $\frac{493}{494}$, $\frac{495}{496}$, $\frac{497}{498}$, $\frac{499}{500}$, $\frac{501}{502}$, $\frac{503}{504}$, $\frac{505}{506}$, $\frac{507}{508}$, $\frac{509}{510}$, $\frac{511}{512}$, $\frac{513}{514}$, $\frac{515}{516}$, $\frac{517}{518}$, $\frac{519}{520}$, $\frac{521}{522}$, $\frac{523}{524}$, $\frac{525}{526}$, $\frac{527}{528}$, $\frac{529}{530}$, $\frac{531}{532}$, $\frac{533}{534}$, $\frac{535}{536}$, $\frac{537}{538}$, $\frac{539}{540}$, $\frac{541}{542}$, $\frac{543}{544}$, $\frac{545}{546}$, $\frac{547}{548}$, $\frac{549}{550}$, $\frac{551}{552}$, $\frac{553}{554}$, $\frac{555}{556}$, $\frac{557}{5$

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:**

```

    /cgn2_6/ptodata/1/iaa/5A_COMB.pcp.*
    /cgn2_6/ptodata/1/iaa/5B_COMB.pcp.*
    /cgn2_6/ptodata/1/iaa/6A_COMB.pcp.*
    /cgn2_6/ptodata/1/iaa/6B_COMB.pcp.*
    /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*
    /cgn2_6/ptodata/1/iaa/backfiles2.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	No.	Score	Match Length		
1	210	6.7	533	US-09-385-259-2	Sequence 2, Appli
2	210	6.7	533	US-09-645-370-2	Sequence 2, Appli
3	209	6.6	533	US-08-488-305A-6	Sequence 6, Appli
4	123.5	3.9	1043	US-08-928-361B-30	Sequence 30, Appli
5	123.5	3.9	1721	US-08-700-651-5	Sequence 5, Appli
6	123.5	3.9	1721	US-08-928-361B-6	Sequence 6, Appli
7	116.5	3.7	1042	US-08-928-361B-11	Sequence 11, Appli
8	116.5	3.7	1837	US-08-928-361B-5	Sequence 5, Appli
9	106	3.4	412	US-09-461-697-77	Sequence 77, Appli
10	104.5	3.3	2172	US-08-611-107-31	Sequence 31, Appli
11	98	3.1	741	US-09-001-984C-106	Sequence 106, Appli
12	97.5	3.1	10182	US-09-134-001C-3159	Sequence 3159, Appli
13	97	3.1	516	US-08-097-829-4	Sequence 4, Appli
14	97	3.1	516	US-08-577-403-4	Sequence 4, Appli
15	96	3.0	1410	US-09-335-409-3	Sequence 3, Appli
16	96	3.0	1410	US-09-568-102-3	Sequence 3, Appli
17	96	3.0	1410	US-09-567-959-3	Sequence 3, Appli
18	96	3.0	1410	US-08-568-480-3	Sequence 3, Appli
19	96	3.0	1410	US-09-568-486-3	Sequence 3, Appli
20	96	3.0	1410	US-09-568-472-3	Sequence 3, Appli
21	96	3.0	1410	US-09-567-899-3	Sequence 3, Appli
22	95.5	3.0	1802	US-09-322-478-18	Sequence 18, Appli
23	94.5	3.0	620	US-09-000-145-5	Sequence 5, Appli
24	94	3.0	1471	US-08-755-587-188	Sequence 188, Appli
25	92.5	2.9	868	US-08-907-166-2	Sequence 2, Appli
26	92.5	2.9	1057	US-08-697-367-23	Sequence 23, Appli
27	92	2.9	205	US-09-385-259-3	Sequence 3, Appli

28	2.9	205	4	US-03-645-370-3	Sequence 3, Appl
29	2.9	230	4	US-03-625-188-18	Sequence 18, Appl
30	91	733	3	US-08-725-459B-22	Sequence 22, Appl
31	90.5	470	1	US-08-090-523-10	Sequence 10, Appl
32	90.5	470	1	US-08-398-627-10	Sequence 10, Appl
33	90.5	470	1	US-08-406-958-10	Sequence 10, Appl
34	90.5	470	5	PCT-US91-04035-10	Sequence 10, Appl
35	90.5	470	5	PCT-US94-05275-10	Sequence 10, Appl
36	90.5	1038	4	US-08-334-179A-2	Sequence 2, Appl
37	90	390	4	US-08-861-774B-18	Sequence 18, Appl
38	90	844	3	US-09-029-267-20	Sequence 20, Appl
39	90	1253	1	US-08-232-966B-12	Sequence 12, Appl
40	90	1261	1	US-08-252-966B-18	Sequence 18, Appl
41	88.5	383	4	US-09-530-212A-2	Sequence 2, Appl
42	88.5	462	1	US-08-548-059-5	Sequence 5, Appl
43	88.5	472	2	US-08-844-057-2	Sequence 2, Appl
44	88.5	872	4	US-09-006-730-2	Sequence 2, Appl
45	88	639	4	US-09-347-801-17	Sequence 17, Appl

ALIGNMENTS

```

RESULT 1
US-09-385-259-2
; Sequence 2, Application US/09385259
; Patent No. 6201114
; GENERAL INFORMATION:
; APPLICANT: Aguirre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; APPLICANT: Ray, Kunal
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
; DATE OF INVENTION: IN DOGS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/385,259
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 60/103,219
; EARLIER FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-385-259-2

```

[illegible]

```

QY 425 VVIGSCMTTPTDSIFNED-ENLKSVLSEIRLNKLTGSTRRPPIISNEDQOVNLEAGWNR 483
Db 325 IVDLCCKWGFEEFYNYLYLANRENWEEVK---KNARKAPQPEVRSVLPNIDKADTGK 381
QY 484 NMLGRKTKFAYLAL-----AEPWPKVSG-----FAKVDLTT--GEVKKHLYG-----523
Db 382 NLVTLPTNTATATLRSDETILWEPEVLFSGPROAFEPQINQYKYGKPYTYAYGLNLH 441
QY 524 -----DNRYGGEPLFLP-GEGGEDEGYILCFVHDEKTKW--SE 559
Db 442 FVDRCKLNKVKETWVWQEPDSYSEPIFVSHPDALDEDDGVLVSVVSPGAGOKPAY 501
QY 560 LQTVNAVSELEATVKLPSRPVPGFHGT 588
Db 502 LLIINAKDLSEVARAEVINIPVTFHGLP 530

RESULT 2
US-09-645-370-2
; Sequence 2, Application US/09645370
; Patent No. 6428958
; GENERAL INFORMATION:
; APPLICANT: Aguirre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
; TITLE OF INVENTION: IN DOGS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/645,370
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/385,259
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: 60/103,219
; PRIOR FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-645-370-2

Query Match 6.7%; Score 210; DB 4; Length 533;
Best Local Similarity 19.9%; Pred. No. 5.4e-13;
Matches 113; Conservative 98; Mismatches 218; Indels 140; Gaps 25;

QY 120 SVQI---AGNFAPVNEQPVRRNLP---VVGKLPDSITKGVYVNGANPLH---EPVTGHH 169
Db 2 SIOVEHPAGYKKLFETVEELSSPLTAHVTRGRIPLWLTGSLRCGGLFVGVSEPF--YH 59
QY 170 FPDGGMVHAVKFEHGSASYACRFTQTNRFVQ-----RQLGRPVFPKATGELHGH 221
Db 60 LFDGQALLHKFDEKEGHVTHRRFIRTDAYVRAMTEKRIVITBFGTCAPDPCKNI---115
QY 222 GIARLMLFYARAAGIVDPAGHCTGVANAGLYVFNGLLANSDE-----DLPYQVQITPNG 276
Db 116 -FRFSYSF-----RGVEVDNALV-----NVPVGEDYACTEINFITKINPE-158
QY 277 DLKTVGRFDPDGOLE-STMLAHKVPDPES-----GELFALSVDVSKPVLYKVPRES 326
Db 159 TLETIKOVDLCNVSVNGATAHPIENDGIVYVINGCNFCGKNFSIAYNIVKIPLOADKED 218
QY 327 PDGKSPDVEIQLD-----OPTMMHDAITENFVVVDPQOVFKLPPEMIRGGS-----374
Db 219 P--ISKSEVVVVQPCSDRRFSPVSHSGLTPNVIVFVETPVKINLLKFLSWSLWGANYM 276
QY 375 -----PVVYDKNKVARFGTLDKXAESSNKKWIDAPDFCFHLNANWEPETDEV 424
Db 277 DCFESNETMGVHLIADKKRKKVYLNKKYRTSSFNL-----PFIINTYEDNEF--L 324
QY 425 VVIGSCMTTPTDSIFNED-ENLKSVLSEIRLNKLTGSTRRPPIISNEDQOVNLEAGWNR 483

```

```

Db 325 IVDLCCKWGFEEFYNYLYLANRENWEEVK---KNARKAPQPEVRSVLPNIDKADTGK 381
QY 484 NMLGRKTKFAYLAL-----AEPWPKVSG-----FAKVDLTT--GEVKKHLYG-----523
Db 382 NLVTLPTNTATATLRSDETILWEPEVLFSGPROAFEPQINQYKYGKPYTYAYGLNLH 441
QY 524 -----DNRYGGEPLFLP-GEGGEDEGYILCFVHDEKTKW--SE 559
Db 442 FVDRCKLNKVKETWVWQEPDSYSEPIFVSHPDALDEDDGVLVSVVSPGAGOKPAY 501
QY 560 LQTVNAVSELEATVKLPSRPVPGFHGT 588
Db 502 LLIINAKDLSEVARAEVINIPVTFHGLP 530

RESULT 3
US-08-488-305A-6
; Sequence 6, Application US/08488305A
; Patent No. 5679772
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof; Eriksson, Ulf; Peterson, Per A.
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which
; TITLE OF INVENTION: bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5679772
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,305A
; FILING DATE: 7-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohli, Vineet
; REGISTRATION NUMBER: 37,003
; REFERENCE/DOCKET NUMBER: LUD 5280.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-488-305A-6

Query Match 6.6%; Score 209; DB 1; Length 533;
Best Local Similarity 20.7%; Pred. No. 6.8e-13;
Matches 119; Conservative 99; Mismatches 222; Indels 134; Gaps 29;

QY 105 VSHEKHLPLPTKADPSVOIAGNEAPVNEQPVRRNLP-----VVGKLPDSITKGVYVNGANP 160
Db 1 MSSQVHP-----AGGYKKLFETVEELSSPLTAHVTRGRIPLWLTGSLRCGGL 49
QY 161 LH---EPVTGHHFDPDGMVHAVKFEHGSASYACRFTQTNRFVQ-----RQLGRPV 209
Db 50 FEVCSSEPF--YHLDGQALLHKFDEKEGHVTHRRFIRTDAYVRAMTEKRIVITBFGTCA 107
QY 210 FPKAIGELGHGTGIARLMLFYARAAGIVDPAGHCTGVANAGLYVFNGLLANSDE-----264
Db 108 FPDPCCKNI-----FSRFSYSF-----RGVEVDNALV-----NIYPVGEDYACT 147
QY 265 DLPYQVQITPNGDLKTVGRFDPDGOLE-STMLAHKVPDPES-----GELFALSVDV 314

```

Db 148 ETNFKVNEP-TLETIKQVLDLCNVSVNGATAHPHIENDGTVYVNIQNCFCGKNEFSIAYNI 206
QY 315 VSKPYLYFFSPDGTSPQVEIQLD-----QPTMMHDFALTENFVVPVQVVFVKLP 369
Db 207 VKIPPLADKEDP--ISKSEIVVQFPCSDREKPSYVHSFGLTPNYIVFVETPVKINLFKE 264
QY 370 IRGSPVYVYDKNKVARF-----GILDKYAEDSSNIKWIDAPOCFCFHLNANWEEP 419
Db 265 LSSWS--LWGANYMDCFESNETMGWLHIAKKRKKYINNRYTSP-FNLFHHINTYEDH 321
QY 420 ETDEWVIGSOMTPDPSIFNESD-ENLKSVLSEIRLNKKTGSTRPLIISNEDQOVNLEA 478
Db 322 EF--LIVDLCCWKGFEFVYVILYLANLRENWEEK---KNARKAPQEVRYVPLNLDK 376
QY 479 GVMNRNM--LGRKTKFAYALAEF-W--PKV--SG-----FAKVDLT--GEVKKHLYG 523
Db 377 ADTCKNLVTLPNTTATILCSDETILWLEPEVLFSGPRGAEFPQINCKYGGKPYTVAYG 436
QY 524 -----DNRYGGELFLP--GEGGEDEGYILCFVHDEKTV 556
Db 437 LGLHNFVDPRLCKLNVRTKETWVQEPDSYSEPIFVSHPDALBEDDGVLSVYVSPGAG 496
QY 557 K--SELOIVNAVSVLEVEATVKLPSRVPGHGT 588
Db 497 QKPAYLILNAKDLSEVARAEVINIPVTFHGLF 530

RESULT 4

US-08-928-361B-30
; Sequence 30, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verry, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1043 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-30

Query Match 3.9%; Score 123.5; DB 3; Length 1043;

Best Local Similarity 20.6%; Pred. No. 0.0021;
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

QY 5 TATAAAGSGLGGNHTQPLPSSOSSDLISYCSLL--PMASRVTRK-----LNVSSALHTPP 58
Db 153 TIAGIVSG-----ISASESL-LSQKSALIDATNNVVGEEFGLLNPAIGVMIPG 200
QY 59 ALHPFKGSSNSPAI-----VVKPKAKESNTKOMNL-----FORAAALD 98
Db 201 FLGSEQIOFSPEDGGIIPPEVAANADKFLSIPSPESIPKDKIDISISEMYD 260
QY 99 AAGFLVSHKHLPLPKTADSPVOIAGFAPVNEQPVRRNLPPVVGKLPDSIKGVYVRGA 158
Db 261 IESGRLLGQVSKRPIGSG-----IAGLNPIMKTIPTQD--SVTKGPIIDPTIGL----- 307
QY 159 NPLHEPVTHGHFFDGDGMVHAVKEFHGSASY--ACRFTQTNRFFVQEROLGRPFVFKAI 216
Db 308 -PFNPP-TGH-----LINTNNNTMDSFAGAYKAYAVSNGIKTDNVYGLPV-----GE 353
QY 217 LHGTGTIARLMLFYARAAAGIVDPAHGTGVAN--AGLIVFNGRL-LAMSEDDLPKQVOIT 273
Db 354 ITGLPKDPGSDIPFNSTTGTGELVDPSTGKPIINNSTAGIV--SKPKGLPIEDE----- 403
QY 274 PNGDLKIVGRED-----FDGQLESTMIAPKPVDPESGELFALSYDVVSKPKLYFRFSP 327
Db 404 -NGNL-----FDPSTNLPIDGN-----NOLVNPENSTVSGSTSGTTP-----KP 443
QY 328 -----DGTKSPDVEI--QLDQ-----PTMMDFAITENFVVPVQVVF 364
Db 444 GIPVNGGVVPEEAKQADKQKGLIIVPTNSINKDPVTNTQYSNTTNGIINPETGKV- 502
QY 365 KLPENIRG-----GSPVYDKNKVARFGILDKVAEDSSNIKWIDAPDCFCFHLNWE 417
Db 503 -IPGSLGPLNYPSENTPQQTDE-----ITGKPDVTVTGLPY----- 538
QY 418 EPEDEVVVGSCMPPPPSIENSDENLKSVLSEIRLNKKTGSTRPRIISNEDQOVNLE 477
Db 539 DESTEIIDPAIKLPIGSAVD-----EILTEV-LAITDVTGTLFI-----DLE 583
QY 478 AGMVNRNMLGRKIKFAYLALAEPPKVSFAKV---DLTGTGVKKHLYGDNRYGGEPLFL 534
Db 584 TGL-----PRDPVSGLPOLPNTGTLVDPNSKKKPIPGS--HSG---FI 619
QY 535 PEGGSEE 541
Db 620 NGTSGEQ 626

RESULT 5

US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF CRYPTOSPORIDIUM PARVUM
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum

RESULT 7

US-08-928-361B-11
; Sequence 11, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1042 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-11

Query Match 3.7%; Score 116.5; DB 3; Length 1042;
Best Local Similarity 19.9%; Pred. No. 0.01;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHPKQSSNSPAI-----VVKPKAKESNTKOMNL----- 88
Db 190 LNPATGVMTIPGSLGSPSEQTPFSEIDGGIIPPEVAANADKFKLSIPSPVESPEKDO 249
QY 89 -PQRAAALDAEGLVSHKHLPLKPTADPSVQIAGNFAVPNEQPVRRNLPVVGKLPD 147
Db 250 KIDSISELMYDIESGLICQVSKRPIPGS-----IAGSLNFMKPTPTD-SVUGKPID 302
QY 148 STKGVVVRNGANPLHEPTVGHFFDGDGMVHAKVEHGSASY--ACRFTQTRFVOEROL 205
Db 303 PTTGL-----PFNPP-TGH-----LINPTNNTMDSFAGAYKAVSNGIKTDNVY 347
QY 206 GRPV-----FPKAIHELHGTGIAIPLMFLYARAAGIVDPAGHTGVAN-- 248
Db 348 GLPVDEITGLPKDPVSDIPFNSTTGL-----VDFSTGKPINNYT 387
QY 249 AGLVFNQRLIAMSDDLPIYQVITPNGDL---KTVGRFDFDQLESTMIAPKVDPESG 305
Db 388 AGIV-SGRGLPIEDE-----NGNLPDPSTKLPIDGNQL-----VAPETN 428
QY 306 ELFALSIVDVSKPYLYKFRFSP-----DGTKSPDVEI--QIDQ----- 342
Db 429 STVSGTSGSTKP-----KPIGVNGGVVPEDEAKDQADKDGKDLIVPTNTSINKDP 481

QY 343 TMMHDFAITENFVVVQOVVEKLEPMIRG-----GSPVVYDKNKVAREGILDKYAED 395
Db 482 VTNTQSYNTTIGNINPETGAV--IPGSLPGSLNVPSENTPQOTDE-----ITGRPVDT 532
QY 396 SSNIKWIDAPDCFCFHLNNAWEPEPDEVVVGSCMTTPPDSIFNESDENLKSVLSEIRLN 455
Db 533 VTGLPY-----DPSTGEIIDPATKLPDIPGSVAGD-----EILTEV-LN 569
QY 456 LKGTGSTRRLIISNEDQOVNLEAGMVRNMLGRKTFAYLALAEPPKPVSGFAKV----DL 512
Db 570 ITTDEVITGLPI-----DLETGL-----PRDPVSGLPQLPNGTL 602
QY 513 TTGEVKKHLYDNRYGGEPLFLPGEEGEE 541
Db 603 VDPSENKKPIPGS--HSG---FINGTSGEQ 626

RESULT 8

US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-5

Query Match 3.7%; Score 116.5; DB 3; Length 1837;
Best Local Similarity 19.9%; Pred. No. 0.03;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHPKQSSNSPAI-----VVKPKAKESNTKOMNL----- 88
Db 985 LNPATGVMTIPGSLGSPSEQTPFSEIDGGIIPPEVAANADKFKLSIPSPVESPEKDO 1044
QY 89 -PQRAAALDAEGLVSHKHLPLKPTADPSVQIAGNFAVPNEQPVRRNLPVVGKLPD 147

Db 1045 KIDSISELMYDIESGRILGOVSKRPIGS-----IAGDLNPMKPTQTD-SVIGKPID 1097
QY 148 SIKGVYVNGANLPHVPTGHHFDDGCMYHVKFEGSASY--ACRFOTNRFVQEROL 205
Db 1098 PTTGL-----FNPBP-TGH-----LINPTNNMTDSFAGAYKAYVNSGKIDNVY 1142
QY 206 GRPV-----BPKAIGELHGHGTGIARLMLFYARAAAGVDPHAGTGVAN-- 248
Db 1143 GLPVEITGLPKDPVSDIPNSITGEL-----VDSTGKPINNYT 1182
QY 249 AGLVYFNGRLANSEDDLPVOVQITPNDL---KTVGRFDDQOLESTMIAPKVDPESG 305
Db 1183 AGIV-SGKRLGPPIEDE-----NGNLFDPSTKLPIDGNLQ-----VNPETN 1223
QY 306 ELEFALSVDVYKPKLYKFRESP-----DGTKSPDVEI--OLDO-----P 342
Db 1224 STVSGSTSGTKP-----KPGIPVNGGVDPDEAKQADKDGKGLIIVPPNSINKDP 1276
QY 343 TMMHDAITENFVVVDQOVVFKLPEMIRG-----GSPVYVDKNKVARFGILDKYAE 395
Db 1277 VNTQYSNTGNTGINPETGKV--IPGSLPGSLNKPSEFNTPOQDE-----ITGRPVDI 1327
QY 396 SSNIKWIDAPDCFEHLNWAWEPEDEVVVIGSCMTPPDSIFNESDENLKSVLSEIRLN 455
Db 1328 VTGLPY-----DFTCEIILDPAKLPVPGSVAGD-----EILTEV-LN 1364
QY 456 LKGTGSTRPILISNEDQQVNLNLAGVMNRNMLGRKTRKFAYLALAEPPKVSQPAKV--DL 512
Db 1365 ITTDEVTGLPI-----DLETL-----PRDPVSLPQLPGL 1397
QY 513 TTCEVKKHLYGDNRYGGEPLFLPGESEE 541
Db 1398 VDPNKKPIPGS--HSG---FINGTSGEQ 1421

RESULT 9
US-09-461-697-77
; Sequence 77, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FASTSEQ for Windows version 4.0
; SEQ ID NO 77
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-77

Query Match 3.4%; Score 106; DB 4; Length 412;
Best Local Similarity 20.6%; Pred. No. 0.027;
Matches 74; Conservative 45; Mismatches 131; Indels 110; Gaps 16;
QY 23 PLSSQSDLSYCSLPMASRTRKLVNSALHTPPALHPKQSSNSPALVVKPKAK--- 79
Db 9 PISSQTSNRSSP-DLLPREFRLVE-----VHDP-LHQPANKPKPTMLDIPSECS 59
QY 80 -----ESNTKONMLFORAAALDAEGFLVSEKHLPLPKPTADPSVQIAGNAPV 130
Db 60 LATHIQLIQHNRRLNRLIATAQAQNOQOQTEG--VKTESEPLP-SCPGSPPLPDDLPL 116

QY 131 N-EQPVARNLPVVGKLPDSTKAGVYVRNGANPLHEPV-----GHHF 170
Db 117 DCKNP---NAPFOIRHSDPESDFYRGKG-----EPVTELSWHSCRQLLYQAVAILAHAG 168
QY 171 FDGDMYHVKFEGHSYACRFTQTNRFVQERQ--LGRPVFPKAIKELHGHGTGIARLML 228
Db 169 FDCANESVLETLIDVAHEYCLKFTKLLRFVADREARLGQTFFPDVMEQVFEHVGIGSVLS 228
QY 229 F-----YARAAAGVDPHAGTGVANAGLVVFNGLLAMSE 263
Db 229 LQKFWOHRIRKIDYHMYLQISKQLESEYER---IVNPEKAT-----E 266
QY 264 DDLPYQVOITPNDGLKTVGRFDDGQLESTMIAPKVDPP-----ESGELFALSVDVYVSKP 318
Db 267 DAKPVKIKKEEFVSDI---TFPVSEELADLASGDQSLPMGLVGAQSERFPSPNLEVEASP 322

RESULT 10
US-08-611-107-31
; Sequence 31, Application US/08611107
; Patent No. 5801233
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING
; TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,107
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 07/956,700
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/09340
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 08/422,560
; FILING DATE: 14-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARCD:221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2172 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-611-107-31

Query Match 3.3%; Score 104.5; DB 1; Length 2172;
Best Local Similarity 19.9%; Pred. No. 0.74;
Matches 118; Conservative 80; Mismatches 247; Indels 147; Gaps 27;

QY 66 SSNSPAIVKPKAKESNTKOMNLFQRAAAALDAAEGFLVSHKHLPLPKTADPSVQIAG 125
Db 1406 TGNCTVDIYREVDPNTHK--LFYRSATPAGLHG--IALHEPKPL-----DAIDLKR 1457
QY 126 NFAPVNPQVRNLPVVGKL-----PDSIKGVYVRNGANPLHBPVTGHHFFDGDGM--- 176
Db 1458 AARKNETTYCYDFPLAFETALKKSWESGISHVAESNEHNRQYAEVTELFIADSTGSGWT 1517
QY 177 -VHAVKEHGSASVACRFTQTNRFVQERQORPVF-----PKAIGELHGT 221
Db 1518 PLVPERPPGSGNFGVAVANMKLSTPEPPGGREIIVVANDVTFKAGSFGPREDAFFDAVT 1577
QY 222 GIA---RIMLFYARAAGIYDPAHGTGVANAGLVFNGRLLAMSEDDLP-----YQVQITP 274
Db 1578 NLACERKIPLIYLSATAGA-----RLGVAEIKACFH---VGVSDDOSPERGFHYILTE 1629
QY 275 NGDLKTVGRDFDQQLSTMAHPKVDPESGELFALSVDVYSK----- 317
Db 1630 Q-----DY-SRLSSVIAHELKVPESGETRWVVDITVGGKEDGLGCENLHSGGAIAS 1679
QY 318 PYLKVER--FSPDGTSPDVEI-----QLDQPMHDFALTENFVVVPPQOV 362
Db 1680 AYSKAYRETTLETVGTAIGIGAYLARLGMRCIORLDQPLITGYSALNKLJL--GREV 1736
QY 363 VFKLPEMIRGSPVYDKNKVARFGILDKYAEDSSNIKWIDAPDCFCFHLNNAWEEPTD 422
Db 1737 YSSOMQL--GGPKMATNGVVHLTVSDDLGVSAILKWL-----SVVPPYVG 1781
QY 423 EVVIGSCMTPPDSTIFNESDENLKSVLSEIRLNKLTGSTRRP--IISNEDQOVNLEAGMVN 482
Db 1782 GLPLVKSLLDPPERAVATYFPE-----SCDARAAICGTDITGGKWLSEMF 1827
QY 483 RNMLGRKTKAYLALAPWPK--VSGPAKVDLTGTGEVKKHLYGDNRYGGGELFLPG---- 536
Db 1828 RE-----SFVTELGWAKTVITGRASAKTAQA---LLDFNR--EEDPLFILANWRG 1875
QY 537 -EGGED--PGYI---LCFVHDEKTKSE--LOIVNAVSLSEVATVKLPSRV 580
Db 1876 FSGQRDLFEGILQAGAMIVENLTKYQPAFYIIPKAGELRGGAWVVVDOSKI 1927

RESULT 11

US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331

; GENERAL INFORMATION:
; APPLICANT: Laal, Suman

; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Bellisle, John T

; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011

; CURRENT APPLICATION NUMBER: US/09/001,984C
; CURRENT FILING DATE: 1997-12-31

; PRIOR FILING DATE: 60/034,003
; PRIOR FILING DATE: 1996-12-31

; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 106
; LENGTH: 741

; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis

US-09-001-984C-106

Query Match 3.1%; Score 98; DB 4; Length 741;
Best Local Similarity 21.7%; Pred. No. 0.53;
Matches 139; Conservative 79; Mismatches 216; Indels 208; Gaps 37;

QY 97 LDRAGFLVSHKHLPLP-----KTADPSVQIAGNFAVNEQPV---RRNLPPV---GK 144

Db 78 MDAYRQELTEIGYLLPEPDFTITSGVDAEITTTAGPQLVWPVNLNARFALNANAGWS 137

QY 145 LPDSIKGVYV---RNGA--NPLPEPVTG-----HFFDGDGKVHAKFEGHSASVACR 192

Db 138 LYDALYGTVDIPEIDGAEKGPYTNKVRGDKVIAYARKFLD-----DSVPLSSGSGDATG 192
QY 193 FTOINRFVQERQLGPPVPPKAI-----CELHGHTGIAR-----LMLFYARAAGVVDPAH 242
Db 193 FT-----VQDQLVVALPDKSTGLANPQOFAGYTGAAESPTSVLLINHLGHEILIDPES 247
QY 243 GTGVAN-AG-----LVFNGRLAMSEDD--LPYQVQITPN-GDLKTV----- 281
Db 248 QVGTIDRAGVADVITLESAITTIMDFEDSVAADAAKVLGYRNWLGKGLDAAAADKDG 307
QY 282 -----GREDFDQ-----LESTIAHPKVPDPESGELFALSVDVVS 316
Db 308 TAFRLVLRNDRNYTAPGGGQFTLPGRSLMFVRNVGHLMNDIAIVTDCGSEVFEGIMDALF 367
QY 317 KPYLKYFRFSPDGTKSPDV-----ETQLDQPTMMH---DFAIT--ENFVAVVPPQOV 362
Db 368 TGLI-----AIHGLKASDVNGPLNSRTGSIYVKP--KMHGPAEVAFTCELSRVED--- 418
QY 363 VFKLPEMIRGSPVYDKNKVARFGILDKYAEDSSNIKWIDAPDCFCFHLNNAWEEPTD 422
Db 419 VLGLPQ-----NTMKIGIMDEERTTVNLK-----ACIKAAAD 451
QY 423 EVVIGSCMTPPDSTIFNESDENLKSVLSEIRLNKLTGSTRRP-----IISNEDQOVN 475
Db 452 RVFINTGF-----LDRTGDEIHTSWEAGPMWRKTKMSQPWILAYEDHNVD 498
QY 476 --LEAGMVNRNMLGR-----KTKFAY--LALAPW-PKVSG-----FAKV 510
Db 499 AGLAAGFSGRAQVKGWMTMTLMDVMVETKIAQPRAGASTAWVPSPATAATLHALHYHQV 558
QY 511 DLTTGVEVKKHLYGDNRYGGEPL-----FLPGEGGEDE-----GYILCFVUDE 553
Db 559 DVAA--VQDGLAGRRRTIEQLLTIPIAKELAWAPDEIREEDVNNCOSILGYVVRWV-DQ 615
QY 554 KTKWSELQIVNAVSL-EVEATVKLPSPRYPGF--HGTFIGAD 592
Db 616 GVGSCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSAD 657

RESULT 12

US-09-134-001C-3159

; Sequence 3159, Application US/09134001C
; Patent No. 6380370

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-11-08

; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159

; LENGTH: 10182
; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 3.1%; Score 97.5; DB 4; Length 10182;
Best Local Similarity 19.4%; Pred. No. 61;
Matches 106; Conservative 72; Mismatches 204; Indels 165; Gaps 26;

QY 13 RWLGGNHTQFPPLSSQSSDLSYCSLPMASRVTRKL-----NVSSALHTPPALHFFKQS 66

Db 673 RWOGAN-----AMNAYQTEELFSQENLQNAARSGRPQIOFLVGFVEDSHHNPETL-LFVN- 726

QY 67 SNSPAIVKPKAK-----ESNTKOMNLFQRAAAALDAAEGFLVSHKHLPLPKT 116

Db 727 -----LVVKKPELKHTEIYLDHNEKQDKRKFVSVKRAGHVFQVMSGLTHTVSGILPYQ 781

QY	117	ADSVQIAGNAPVNEQVRRMLPVVGLKLPQSI---	KGVYVRNGANPLHEPVVTGHHFFDG	173
Db	782	QEIRIKLTN-----EPFKDSEWITGYPNTLQ	NAVGRTNKATEKNLALVGH--IDP	833
QY	174	DGMVHAKVEHGSASACRFTQTRFVQEROLGR	VPFKAJ--GEJLGH-----T	221
Db	834	GNFYITVKECDKVEQFEIR-----SKPTPR	PIIITITANELGRNPNHKEIRVMT	880
QY	222	GTARLMLFYARAAAGIVD-----PAHOTG	VANAGLVYF-----NGRLLAMSE	263
Db	881	DIPNDTTAKIKLVMGCTGDGHDPEINPYTV	PENYTVAAE--VYHNDPSKNGVLFRSS	937
QY	264	D---DLPYQVQITPNCGLTKVCRF-----	DFDGOLE--STMIAHPKVDPSSGELFALS	313
Db	938	DYLUKDLPL-----SGEUKAIVYQYVQSF	NSVPSSDITPTTINEFAGLVH-----	986
QY	314	VWSPKPYLYKFR-----FSPDGTKSPD	VEIQLDQPTMMHDFAITENFVVVPDQVQ	362
Db	987	-----KYRGDHVEIITLPVDNTGSGSLR	DVNVNLPQ-----GWTKFIINPNNTT	1032
QY	363	VFKLPEMIRGGSPVYVQKNKVARFGILD	KYADGSNIIKWIDAPDCFCFHLWNAWEEPT	421
Db	1033	EGTL--KLIGNIPSEAYNTTYHENIT--	ATDMSG-----NTINPAKTE	1072
QY	422	-----DEVVVIG-----SCMTTP	PDPSINESDENLKSVLSEIRLNLKTGSTR	467
Db	1073	ILNVGKLADLNPVLGSLRDLQVLVTP	SSLSNREEVKVKISEANANIRSYLQNNPIL	1132
QY	468	SNEDQVQ	474	
Db	1133	AGVNGDV	1139	

RESULT 13
 US-08-097-829-4
 ; Sequence 4, Application US/08097829
 ; Patent No. 5498831
 ; GENERAL INFORMATION:
 ; APPLICANT: Burgess, Diane G.
 ; APPLICANT: Dooner, Hugo K.
 ; TITLE OF INVENTION: Pea Adp-Glucose Pyrophosphorylase Subunit
 ; TITLE OF INVENTION: Genes and Their Uses
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Kourie and Crew
 ; STREET: One Market Plaza, Steuart Tower, Suite 2000
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94105

```

Query Match          3.1%; Score 97; DB 1; Length 516;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 123; Conservative 69; Mismatches 172; Indels 224; Gaps

QY      1  MASFTATAAVSGRWLGNHTQPPPLSSQSSQSDLSYCSLPMASRVTKLNVSSA-----L 54
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1  MASMAAIGVUK-----VPSSSSSSS-----SSSSSKAIAARNLSFTSSQLCGDKI 46

QY      55  HTPPALHFPKQSSNSPAIVYKPKA-RESNTPQMNLFQRAAAALDAAEGLV---SHEKL 110
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      47  FT---VSGTRRSSGRNPFIVSPKAVSDSKNSQCLOPDASRVL-----GIILGGGAGTSL 99

QY      111  HPL-PKTAADSVQIAGNFADVNEQPVVRNLPVYVKLPDSTIKGYVVRNGANPLHEPVYTHH 169
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      100  YPLTKRAKPAVPLGANYRLI-----DIPVSNCLNSLSKIVYLTQFN---SASLNRH 149

QY      170  F-----FDGDGMVHAVKFE-----HGSA-----SYACRFQTN----- 197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      150  LSRAYASNLGGYKNEGFEVLAQAQQSPENPWFQGTADAVRQYLMWLFEEHNVLVYLVLAG 209

QY      198  ----RFVQERQLGRPPFPKPAIGELHGHGTGIARLMLFYARAAGIVDPAGHTGTVANAGLVY 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      210  DHLRYMDYER-----FIQAHRESADAITVASLPMDEARATA-----FGLMK 250

QY      254  FN-GRLLAMSE-----DD-----LPYQVQITPNC-----D 277
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      251  IDEEGRIVESEKPKGQLKAMKVDTTILGLDERAKEMPY---IASMGIVYVSKHVMLD 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      278  L---KTVGRFDQGLESTMIAPKVPDPESGEL-----FALSYD----- 313
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      308  LLRDKPFGANDFSE-----VIPGATELGLRVQAYLYDGYWEDIGTTEAFYNAML 357

QY      314  -VVSKEP---YLKYRFPSPDGTSPDVEIQIDQPTMMHDAITENFV---VWPDQGV--- 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      358  GITKKVPDPFSFYDRSSPIYTPQ-----RYLPPSKMLDADITDSVIGEGCVKNCKIHHS 412
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      363  VFKLPEMI-----RGSPVYVDKKNVARFGILDKYAE 394
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      413  VGLRSCISGAETIEDTLNCAQYETDADRRLAAKGGVPIGIGKSHIKRAIDDKNAR 472

QY      395  DSSNIKWIDAPDCCFHLNWAEEPE-----TDEVVVGISCMTPPDSI 437
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      473  IGDVVKIINSD-----NVQEAAREIEGYFKSIGIVIVKDALPSGTV 515
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-08-577-403-4
; Sequence 4, Application US/08577403
; Patent No. 5773693
; GENERAL INFORMATION:
; APPLICANT: Burgess, Diane G.
; APPLICANT: Dooner, Hugo K.
; TITLE OF INVENTION: Pea ADP-Glucose Pyrophosphorylase Subunit
; TITLE OF INVENTION: Genes and Their Uses
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/577,403
; FILING DATE: 22-DEC-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/097,829
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 12176-35
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-577-403-4

Query Match 3.1%; Score 97; DB 1; Length 516;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 123; Conservative 69; Mismatches 172; Indels 224; Gaps 32;
QY 1 MASFTATAVSGRWLGNHTQPPLLSSQSSDLSYCSLPLWASRYTKLVYSSA-----L 54
Db 1 MASMAIGVLK-----VPSSSSSSS-----SSSSSKAIARNLSFTSSQLCGDKI 46
QY 55 HTPPALHFPKQSSNSPAIVVVKPKA-KESNTKOMNLFORAAAAALDAAGELV---SHEKL 110
Db 47 FT---VSGTFRSSGRNPFIVSPKAVSDSKNSQICLPDASRVL-----GILGGGAGTR 99
QY 111 HPL-PKTADPSVQIAGNAPVNEQPVRRNLPVVGKLPDSIKGYVVRNGANPLHEPVTHGH 169
Db 100 YPLTKRAKPAVPLGANRYLI-----DIPVSNCLNSISKIYVLTQFN---SASLNHR 149
QY 170 F-----PDGDMGHVAKFE-----HGSA-----SVACRFTQIN----- 197
Db 150 LSRAYASNLGGYKNEGFEVFLAAQQSPFNWFGTADAVRQVLWLFEEHNVILEYLVLAG 209
QY 198 ---RFVQERQLGRVPFPAIGELHGHGTGIARLMIFYARAAAGIVDPHAGTGVANAGLVY 253
Db 210 DHLRYMDYER-----FIQAHRESADITIVASLPMDEARATA-----FGLMK 250
QY 254 FN--GRLLAMSE-----DB-----LPYQVOITPNG-----D 277
Db 251 IDEGRIVFESEKPKGEQLKAMKVDTITILGLDDEKAKEMPY---IASMGIYVYVSKHVMLD 307
QY 278 L---KTVGRFDFGQLESTIMAHPKVDPESEL---FALSXD----- 313
Db 308 LLRDKEFGANDFGE-----VTPGATELGLRVQAYLYGYWEDIGTIEAFYNANL 357
QY 314 -VYVKP---YLYFRFSPDGTSPDVEIQDQPTMMHDAITENFV---VVPDQCV--- 362
Db 358 GITKKVPDPFSFYDRSSPIYTPQ-----RYLPPSKMLDADITDSVIGEGCVIKNKTHHS 412
QY 363 VFKLPEMI-----RGSPVYDKNKVARFQILDKYAE 394
Db 413 VVGLRSCISEGAIETLTMGADYYETDARRFLAAGKGVPIGIGNSHIKRAIDKNAR 472
QY 395 DSSNIKWIDAPDFCFHLNWAEEPE-----TDEVVVIGSCMTPTPDSI 437
Db 473 IGDVXIIINSD-----NVQEAARETEGYFIKSGIVTVIKDALIPSGTV 515

RESULT 15

US-09-335-409-3
Sequence 3, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1410
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-335-409-3

Query Match 3.0%; Score 96; DB 3; Length 1410;
Best Local Similarity 23.2%; Pred. No. 2.7;
Matches 107; Conservative 57; Mismatches 182; Indels 116; Gaps 25;
QY 10 VSRWLGNGNHTQPPLLSSQS--SDLSY-----CSSLPMAKSVTRK-----LNVSALHT 56
Db 645 VSEAGVEGDGDQPPMPFIQTPSDLAYIYTSGSTGLPKGVWIDHRGAVNTILDINERFEI 704
QY 57 PPALHFPKQSSNS-----PAIVVVKPKA-KESNTKOMNLFORAAAA---AL 97
Db 705 GPDRVLALSSLSFDLSYDVVFGILAAAGTIVVDPDASKLRDPAHWAELIEREKVTIVNSV 764
QY 98 DAAEGELVSHEKLHP--LPKTADPSVQIAGNAPV---NEQPVRRNLPV---GKLPDS 148
Db 765 PALMRMLVEHFEGRPOSLSRLSL--LSGDWIPVGLPGLQAIRPGVSVISLGGATEAS 823
QY 149 I--KGIVVRNGANPLHEPVTHGHFFDGDGMVHAKFEGHSASYACRF--TOTNRFVQERQL 205
Db 824 IWSIGYFVRN-----VDLSWASIPYGRPLNQTFFHVLDEALE 860
QY 206 GRPVFPAIGELH-GHTGIARLMIFYAR-----AAAGIVDPAHGTGVANAGLYVFNGLL 259
Db 861 PREVW--VPQOLYIGGVGLA---LGWWRDEEKTARKSLVHPETGERLYKTGDL---GRYL 912
QY 260 AMSEDDLPYQVOITPNGDLAKTVGREFD-----GQLESTIMAHPKVDPESELFA 309
Db 913 -----PGNIEFMGRDNQIKLGYRVSELGEIETLKSHPNW--RDVIVP 956
QY 310 LSYDVVSKPYLYFRFSPDGTSPDVEIQDQPTMMHDEFA--ITENFVVVDPDQCVVFKLP 367
Db 957 VGDAANKLLLAY--VYPEGTRRRRAAQDASLAKTERIDARAHAAEADGLSDGERVQFKLA 1014
QY 368 EM-IR---GGSPVYDKNKVARFQILDKYAE DSSNIKWIDAP 405
Db 1015 RHGLRRDLGKPVVDLTGQDPREAGLDVYARRRSVRTLEAP 1056

Search completed: May 19, 2003, 15:49:59
Job time : 45 secs

BEST AVAILABLE COPY

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:48:44 ; Search time 58 Seconds
(without alignments)
996.104 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVSGRWLGNGHT.....VPYGHCTFIGADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3150	100.0	599	10	US-09-758-269-6
2	2280.5	72.4	605	10	US-09-758-269-16
3	2168.5	68.8	512	10	US-09-758-269-12
4	1991	63.2	593	10	US-09-758-269-2
5	1930	61.3	604	10	US-09-758-269-14
6	1663.5	52.8	577	10	US-09-758-269-10
7	968	30.7	595	10	US-09-758-269-4
8	938	29.8	538	10	US-09-758-269-8
9	937	29.7	538	10	US-09-758-269-18
10	936	29.7	538	10	US-09-758-269-33
11	411.5	13.1	505	1	US-08-976-063C-22
12	260	8.3	526	9	US-10-053-192-1
13	247	7.8	506	9	US-10-053-192-4
14	185.5	5.9	529	9	US-10-053-192-5
15	107	3.4	26926	9	US-09-759-508B-2
16	106.5	3.4	2701	9	US-10-171-311-83
17	106	3.4	412	10	US-09-922-261-77
18	105	3.3	5701	10	US-09-864-761-37319
19	103.5	3.3	1311	9	US-10-103-377C-6

```

20 103 3.3 1156 10 US-09-815-242-13187 Sequence 13187, A
21 101.5 3.2 4342 10 US-09-815-242-5107 Sequence 5107, Ap
22 98.5 3.1 2828 9 US-10-176-847-54 Sequence 54, Appl
23 98.5 3.1 2828 10 US-09-905-129-21 Sequence 21, Appl
24 98.5 3.1 2828 10 US-09-991-630-21 Sequence 21, Appl
25 98 3.1 748 9 US-09-272-975-2 Sequence 2, Appli
26 98 3.1 753 9 US-09-272-975-58 Sequence 58, Appli
27 97.5 3.1 411 9 US-09-738-626-3692 Sequence 3692, Ap
28 97.5 3.1 1302 9 US-09-738-626-5616 Sequence 5616, Ap
29 96 3.0 666 10 US-09-815-242-10929 Sequence 10929, A
30 96 3.0 1410 9 US-10-014-717-3 Sequence 3, Appli
31 95.5 3.0 788 10 US-09-815-242-10562 Sequence 10562, A
32 95.5 3.0 1802 10 US-09-965-553-18 Sequence 18, Appli
33 94 3.0 1080 10 US-09-904-380-2 Sequence 2, Appli
34 94 3.0 1322 9 US-10-083-357-1332 Sequence 1332, A
35 93 3.0 566 10 US-09-815-242-13371 Sequence 13371, A
36 92.5 2.9 868 10 US-09-391-340-2 Sequence 2, Appli
37 92.5 2.9 868 10 US-09-948-369-2 Sequence 2, Appli
38 92.5 2.9 1057 9 US-10-217-700-2 Sequence 2, Appli
39 92.5 2.9 1057 10 US-09-918-909-29 Sequence 29, Appli
40 92.5 2.9 1131 10 US-09-801-368-72 Sequence 72, Appli
41 92 2.9 1167 9 US-10-147-026-8 Sequence 8, Appli
42 91.5 2.9 1081 9 US-10-217-700-4 Sequence 4, Appli
43 90.5 2.9 729 10 US-09-919-497-51 Sequence 51, Appli
44 90.5 2.9 916 10 US-09-745-763-174 Sequence 174, App
45 90.5 2.9 1038 10 US-09-908-500A-2 Sequence 2, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-758-269-6
; Sequence 6, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 599
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-6

Query Match 100.0%; Score 3150; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 3.5e-280;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASFTATAVSGRWLGNGHTOPPLSSQSSDLSYCSSLPMASRVTRKLNYSALHTPPAL 60
Db 1 MASFTATAVSGRWLGNGHTOPPLSSQSSDLSYCSSLPMASRVTRKLNYSALHTPPAL 60
Qy 61 HFPKQSSPAIVVVKPKAKESNTKQMNLFQRAAAALDAAEGFLVSHEKHLPLPKTADPS 120
Db 61 HFPKQSSPAIVVVKPKAKESNTKQMNLFQRAAAALDAAEGFLVSHEKHLPLPKTADPS 120
Qy 121 VQINGNAPVNEQVRNRLPVWGKLPDSIKGVYVYRNCANLHPHPTGHHFFDGDGMVHAV 180
Db 121 VQINGNAPVNEQVRNRLPVWGKLPDSIKGVYVYRNCANLHPHPTGHHFFDGDGMVHAV 180
Qy 181 KFEHGSASYACRFOTNRFVQERQGRPVFPKAIQELGHGTIARLMFYARAAAGIVDP 240

```

```
Db 181 KPEHGSASYACRFTQTNRFVQERQLGRPVPEKKAIGELHGHGTIARLMLFYARAAAGIVDP 240
QY 241 AHGTGVANAGLVFENGELRLAMSEDDLPYQVOITPNGDLKTVGREFDQCOLES 292
Db 241 AHGTGVANAGLVFENGELRLAMSEDDLPYQVOITPNGDLKTVGREFDQCOLES 297
QY 241 AHGTGVANAGLVFENGELRLAMSEDDLPYQVOITPNGDLKTVGREFDQCOLES 352
Db 241 AHGTGVANAGLVFENGELRLAMSEDDLPYQVOITPNGDLKTVGREFDQCOLES 357
QY 301 DPESGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFALTE 352
Db 301 DPESGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFALTE 357
QY 301 DPESGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFALTE 352
Db 301 DPESGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFALTE 357
QY 361 QVFKLPMTIRGSGSPVVDKKNKVARFGLDKYAEADSSNIKWIDAPDCFCFHLNANDEPE 420
Db 361 QVFKLPMTIRGSGSPVVDKKNKVARFGLDKYAEADSSNIKWIDAPDCFCFHLNANDEPE 420
QY 421 TDEVVIGSCMTDPDSIFNESDENLKSLSVLEIRLNKLTGSTRPPIISNEDQOVNLEAGM 480
Db 421 TDEVVIGSCMTDPDSIFNESDENLKSLSVLEIRLNKLTGSTRPPIISNEDQOVNLEAGM 480
QY 481 VNENMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGSGEFLPGE 540
Db 481 VNENMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGSGEFLPGE 540
QY 541 EDEGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRYPYGFHGTGADDLAKQV 599
Db 541 EDEGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRYPYGFHGTGADDLAKQV 599
```

```
RESULT 2
US-09-758-269-16
; Sequence 16, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 16
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-758-269-16
```

```
Query Match 72.4%; Score 2280.5; DB 10; Length 605;
Best Local Similarity 71.7%; Pred. No. 2.1e-200;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;
```

```
QY 5 TATAAAGRWLGNHQPPLSSSQSDLSYCS---SLPMASRVTRKLNVSALHTPEPALH 61
Db 3 TTSHAINTWI-----KTLKMSPEKFEFGFASNSISLLKNQHRQSLNINSSLAQAPILH 57
QY 62 FPKQSSN--SPA--IVVKPKAKESN-----TKOMNLFQRAAAALDAAGFLYSHEKLEH 112
Db 58 FPKQSSNYQTPKNNTISHPKQENNNSSSTSKNVLVORAAAMALDAVESALTKHELEH 117
QY 113 LPKTADPSVQIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGVYVNGANGNLPHEPVTHHFFD 172
Db 118 LPKTADPSVQIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGVYVNGANGNLPHEPVTHHFFD 177
QY 173 GDGMVHAFKFEHGSASYACRFTQTNRFVQERQLGRPVPEKKAIGELHGHGTIARLMLFYAR 232
Db 178 GDGMVHAFKFEHGSASYACRFTQTNRFVQERQLGRPVPEKKAIGELHGHGTIARLMLFYAR 237
```

```
QY 233 AAAGIYDPAHGTGVANAGLVFENGELRLAMSEDDLPYQVOITPNGDLKTVGREFDQCOLES 292
Db 238 GLFGLVDHSGTGVANAGLVFENGELRLAMSEDDLPYQVOITPNGDLKTVGREFDQCOLES 297
QY 293 TMIAHPKLDPVSGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFALTE 352
Db 298 TMIAHPKLDPVSGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFALTE 357
QY 353 NFVVVDDQVQVFKLPMTIRGSGSPVVDKKNKVARFGLDKYAEADSSNIKWIDAPDCFCFHL 412
Db 358 NFVVVDDQVQVFKLPMTIRGSGSPVVDKKNKVARFGLDKYAEADSSNIKWIDAPDCFCFHL 417
QY 413 WNAWEPEPTDEVVIGSCMTDPDSIFNESDENLKSLSVLEIRLNKLTGSTRPPIISNEDQ 472
Db 418 WNAWEPEPTDEVVIGSCMTDPDSIFNESDENLKSLSVLEIRLNKLTGSTRPPIISNEDQ 477
QY 473 QVLEAGMVRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGSGEFL 532
Db 478 QVLEAGMVRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGSGEFL 537
QY 533 FLPGE--GGEDEGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRYPYGFHGTG 590
Db 538 FLPGE--GGEDEGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRYPYGFHGTG 597
QY 591 ADDLAKQ 597
Db 598 ANDLANQ 604
```

```
RESULT 3
US-09-758-269-12
; Sequence 12, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Vigna unguiculata
US-09-758-269-12
```

```
Query Match 68.8%; Score 2168.5; DB 10; Length 612;
Best Local Similarity 68.6%; Pred. No. 4e-190;
Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;
```

```
QY 8 AAVSGRWLGNHQPPLSSSQSDLSYCSLPMASRVTRKLN--VSSALHTPPALHPKQ-- 65
Db 3 SSASNTWFAATLSPFPKLPSTIS--SPTNLLPL--RKTSSNTITCSLOT---LHPKQY 56
QY 66 ----SSNSFAIVVVKPKAKESNT-----KOMNLFQRAAAALDAAL 100
Db 57 QPTSTSTATTTPPTPIKTTTITTTTPPTNPLSDTNPQLPKQKNFLQKAAATDLV 116
QY 101 EGFELVSHKHLPLPKTADPSVQIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGVYVNGANG 160
Db 117 ETALVSHKHLPLPKTADPSVQIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGVYVNGANG 176
QY 161 LHPVTVTHHFFDGDGMVHAFKFEHGSASYACRFTQTNRFVQERQLGRPVPEKKAIGELHGH 220
Db 177 LHPVTVTHHFFDGDGMVHAFKFEHGSASYACRFTQTNRFVQERQLGRPVPEKKAIGELHGH 236
```


Db 252 LVTFNRLAMSEDDLPYHVRVADGDETVGYRDFDGLCCAMIAHAPKLDPAITGELHAL 311
QY 311 SYDWKSPYLKYFRSPDGTGKSPDVEIQLDQPTMMHDFAITENFVVVVDQVVKLPPEMI 370
Db 312 SYDVIRPYLKYFRPDGTGKSDVETPLQPTMIHDFAITENFVVVVDQVVKLPQEML 371
QY 371 RGSPPVVDKKNVARGILDKYAEDSSNIKNIDAPDCFCFHLNANWEEPTDEVVVIGSC 430
Db 372 RGSPPVVDKKNVARGILDKYAEDSSNIKNIDAPDCFCFHLNANWEEPTDEVVVIGSC 431
QY 431 MTPDPSIFNSDENLKSLSVSEIRLNKLGESTRRPIISNEDQOVNLEAGMVRNMLGRKI 490
Db 432 MTPDPSIFNSDENLKSLSVSEIRLNKLGESTRRPIISNEDQOVNLEAGMVRNMLGRKI 490
QY 491 KFYALAEPPKVSFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGE-----EDEGY 545
Db 491 RYALAEPPKVSFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGE-----EDEGY 545
QY 546 ILCFVHDEKTKWSELOTVNAVSLVETATVKLPSRPVYGFHGTFTIGADDLAKQ 597
Db 551 VLFVHDERAGTSELLVYVNAADIRLEATVQLPSRPVYGFHGTFTIGADDLAKQ 602

RESULT 6

US-09-758-269-10

; Sequence 10, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 10

; LENGTH: 577

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-10

Query Match 52.8%; Score 1663.5; DB 10; Length 577;
Best Local Similarity 55.4%; Pred. No. 7.8e-144;
Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

QY 24 LSSQSDLSYCSLIPMASRVTRKLVNVS-ALHTPPALHPFKQSSNSPAIVVVKPKAKESN 82
Db 9 LIPKTPSPRSHLLPOPKNANISRRILLNPFKIPLDITSPVSP-----VKLQPTYPN 62
QY 83 TKOMLFORAAAAALDAE-GFLYSHEKHLPLKPTADPSVOIAGNFAPVNPQVRRNLPV 141
Db 63 ---LNLQKLAATMLDKTSSIVIPMEQNRPLPKPTDPAVOLSGNFAPVNECPVQNGLEV 119
QY 142 VKGLPDSITKGYVVRNANPLHPVPTGHHFFDGDGMVHAVK--FEHGSASYACRFTQTNRF 199
Db 120 VQGISCLKGVYIRNANPMFPPLAGHHLFDGDMHIAVSGFDN-QVSYSCRVTKTNRL 178
QY 200 VOEROLGRVPFPAKIGELHGHGTGIALMLFYARAAAGIVDPAHGTGVANAGLVYFNGRL 259
Db 179 VOETALGRVPFPAKIGELHGHGTGIALMLFYARAGIGLVDTGRGMVANAGVVFNGRL 238
QY 260 AMSEDDLPYQVQITPBGDLKTVGREFDQGLQESTMIAHPKVDPSGELFALSVDVSKPY 319
Db 239 AMSEDDLPYQVQITPBGDLKTVGREFDQGLQESTMIAHPKVDPSGELFALSVDVSKPY 319

QY 320 LKYREFSPDGTGKSPDVEIQLDQPTMMHDFAITENFVVVVDQVVKLPPEMIRGSSPVVD 379
Db 299 LRYLKCTGCKKTRDVEITLPEFTMIHDFAITENFVVVVDQVVKLPPEMIRGSSPVVD 358
QY 380 KKNVARGILDKYAEDSSNIKNIDAPDCFCFHLNANWEEPTDE-----VVVIGSCTMPPD 435
Db 359 KKNVARGILDKYAEDSSNIKNIDAPDCFCFHLNANWEEPTDE-----VVVIGSCTMPPD 417
QY 436 STINESDENLKSLSVSEIRLNKLGESTRRPIISNEDQOVNLEAGMVRNMLGRKTFAYL 495
Db 418 TIFSEGEPTFVSELSEIRLNKLGESTRRPIISNEDQOVNLEAGMVRNMLGRKTFAYL 473
QY 496 ALAEPMPKVSFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGEEDGYILCFVHDEKT 555
Db 474 AIADPPKPCSGIAKVDIQNGTVSEFNYPGSGEPCFVPEGEEDKGYVMGFVRDEK 533
QY 556 WKSELQITVNAVSLVETATVKLPSRPVYGFHGTFTIGADDLAKQV 598
Db 534 DESEFVVVDATDMKQVAAVRLPERVYGFHGTFTVSENLKEQV 576

RESULT 7

US-09-758-269-4

; Sequence 4, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 595

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-4

Query Match 30.7%; Score 968; DB 10; Length 595;
Best Local Similarity 36.7%; Pred. No. 5.3e-80;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLVNVSALHTPPALHPFKQSSNSPAIVVVKPKAKESNTKOMLFORAAAAALDAE 101
Db 42 SPITNPSDNNRRNKKPKTLH---NRTNHTLVSSPPKLRPEMTALALF---TTVEDVIN 94
QY 102 GFLYSHEKHLPLKPTADPSVOIAGNFAPVNPQVRRNLPV-GKLPDSIKGYVVRNANP 160
Db 95 TFIDP-----FSRSVDPKHVLSONFAVLDLPDTCDEIHTGLPLSLNAYIRNGENP 149
QY 161 LHPEVTGHHFFDGDGMVHAVKFEHGSASYACRFTQTNRFVQEROLGRVPFPAKIGELHGH 220
Db 150 QFLPRGPHVLEFDGDMHIAKIHNGKATLCRSRYKTYKYNVEKOTGAPVMNPFSENGV 209
QY 221 T-GTARLMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPBGDLK 279
Db 210 TASVARGALTAARVLITQYNFNVNGIGLANTSLAFFSNLFALGESDLPYAVRLTESGDIE 269
QY 280 TVGREDDQGLQESTMIAHPKVDPSGELFALSVDVSKPYLKYREFSPDGTGKSPDVEI-Q 338
Db 270 TIGRTDFDQGLKAMSTAHPKTDPIITGETAFRYGPV-PPFLTYFEDSAGKKQDVPIFS 328
QY 339 LDQPTMHDFAITENFVVVVDQVVKF---LPMEIRGSSPVVYDKNKVARFGLDKYAE 395
Db 329 MTPSPFLHDFAITKKAIFAETQLGMRNMLDLVLLEGSGPVGTGDKGKTPRLGVIPKYAGD 388

Db 344 EDEVVLIITCRLENPLDMVSGKVKLENFNGNELYEMRNMKTGSASOKKLASASA----- 398
Qy 476 LEAGMVNRMGRKTKFAYLALAEWPVKVSGFAKVD-----LTTGEVKKHLY-- 522
Db 399 VDFPRINCYTGKKORYVYGTTILDSIAKVIKILFEDLHAEATGKRMLEVGNGIKGIYDL 458
Qy 523 GNRVGGEPFLFPGGGREDEGYILCFVHDEKTSKSELIQIVNAVSLVB--ATVKLPSRV 580
Db 459 GEGRGSEALYVPRETAEDDGLYFFVHDENTGKSCVTVIDAKTMSAPVAVVLPKRV 518
Qy 581 PYGFHGTFTGADDLAKQVV 599
Db 519 PYGPHALFVTEEQLOEOTL 537
RESULT 9
US-09-758-269-18
; Sequence 18, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-18
Query Match 29.7%; Score 937; DB 10; Length 538;
Best Local Similarity 37.0%; Pred. No. 3,1e-77;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;
Qy 64 KOSSNSPAIVVKKPK-AKESNTKQNMFLFORAAAAALDAEGLVSHKELHLPKTAADPSVQ 122
Db 4 KLSGSSIIIVHPRPSKGFSSKLLDLRLVVKLM-----HDASLPLH-----Y 47
Qy 123 IAGNFAPV-NEQPVRRNLPPVVKLPDSTIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAYK 181
Db 48 LSGNFAPRDETPPVVKLPDSTIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAYK 107
Qy 182 FEHGSASYACRFTQNRVQERQLGRPVFPKAIAGELHGTGTIARLMLFYARAAAGIVDPA 241
Db 108 IKDGKATVSRVYKTSRLKQEEFFCAAKFMK-IGDLKFFGGLLMVYQOOLRKLKLDNT 166
Qy 242 HGTGVANAGLVFNGRLLAMSEDDLPYQVQITPNDGLKTVGRFDGQLESTMIAPKVD 301
Db 167 YONGTANTALVYHGGKLLAQEADRPYIKVLEDDGLQTLGIIDYDKRLTHSFTAPKVD 226
Qy 302 PESGELFALSYDVSKPYLKYRFPSPDGTSPDVELQDQPTMHDFATENFVVVDDQ 361
Db 227 PVTGEMFTGYS-HTPPYLYTVRISKDGLMHDPVPTITISEPTMHDFATETAYAFMDLP 285
Qy 362 VYFKLPEMIRGSGSPV-YDNKVARFGLDKYAEDESSNKKWIDAPDCFCFLHNAWEEPE 420
Db 286 MHFRPKEMVKKMIYSFDPKAREGLVPYAKOELMIRWELPNCFLFHNANAWEE-E 343
Qy 421 TDEVVIGSCMTTPD-----SIFNESDENLKSVLISIRLNKLTGSTRRPIISNEDQOVN 475
Db 344 EDEVVLIITCRLENPLDMVSGKVKLENFNGNELYEMRNMKTGSASOKKLASASA----- 398
Qy 476 LEAGMVNRMGRKTKFAYLALAEWPVKVSGFAKVD-----LTTGEVKKHLY-- 522

Qy 396 SSNIKWIDAPDCFCFLHNAWEEPEDETVVIGSCMTTPDSTIFNESDENLKSVLISIRLN 455
Db 399 ESEKWFEPVGNIIHAINAWCEDDNGSVVLIAPNIMSIEHTLERMD-LVHALVEKVKID 447
Qy 456 IATGSTRPIISNEDQOVNLEAGMVNRMGRKTKFAYLALAEWPVKVSGFAKVDLITG 515
Db 448 LVTGIVRRHPISAR-----NLDFAVINPFLGRCRSRYVAAAGDPMKISGVVVKLVDVSKG 502
Qy 516 E-----VKKHLGDNRYGGEPLFLPQEGG-----BEDEGYILCFVHDEKTSKSELIQIVNAV 566
Db 503 DRDDCTVARMYSGCYGGEPPFVARDPQNPABEDDGYVVIYVHDEVGTGSKELVMDAK 562
Qy 567 S--LEVEATVKLPSRVPGFHTGFIADDLAK 596
Db 563 SPELEIVAARLPRPVYGFHGLFVKESDLNK 594
RESULT 8
US-09-758-269-8
; Sequence 8, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-8
Query Match 29.8%; Score 938; DB 10; Length 538;
Best Local Similarity 37.0%; Pred. No. 2.5e-77;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;
Qy 64 KOSSNSPAIVVKKPK-AKESNTKQNMFLFORAAAAALDAEGLVSHKELHLPKTAADPSVQ 122
Db 4 KLSGSSIIIVHPRPSKGFSSKLLDLRLVVKLM-----HDASLPLH-----Y 47
Qy 123 IAGNFAPV-NEQPVRRNLPPVVKLPDSTIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAYK 181
Db 48 LSGNFAPRDETPPVVKLPDSTIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAYK 107
Qy 182 FEHGSASYACRFTQNRVQERQLGRPVFPKAIAGELHGTGTIARLMLFYARAAAGIVDPA 241
Db 108 IKDGKATVSRVYKTSRLKQEEFFCAAKFMK-IGDLKFFGGLLMVYQOOLRKLKLDNT 166
Qy 242 HGTGVANAGLVFNGRLLAMSEDDLPYQVQITPNDGLKTVGRFDGQLESTMIAPKVD 301
Db 167 YONGTANTALVYHGGKLLAQEADRPYIKVLEDDGLQTLGIIDYDKRLTHSFTAPKVD 226
Qy 302 PESGELFALSYDVSKPYLKYRFPSPDGTSPDVELQDQPTMHDFATENFVVVDDQ 361
Db 227 PVTGEMFTGYS-HTPPYLYTVRISKDGLMHDPVPTITISEPTMHDFATETAYAFMDLP 285
Qy 362 VYFKLPEMIRGSGSPV-YDNKVARFGLDKYAEDESSNKKWIDAPDCFCFLHNAWEEPE 420
Db 286 MHFRPKEMVKKMIYSFDPKAREGLVPYAKOELMIRWELPNCFLFHNANAWEE-E 343
Qy 421 TDEVVIGSCMTTPD-----SIFNESDENLKSVLISIRLNKLTGSTRRPIISNEDQOVN 475

Db 399 VDFPRINECYTGCKQRYVYVGTILDSIAKVTGIKFDLHAEATGKRMLEVGNGIKGIYDL 458
QY 523 GDNRYGGEPLFLPGEGGEDEGYILCFVHDEKTKWSELOIVNAVSLVE--ATVKLPSRV 580
Db 459 GGRYGSEAIYVPRETAEDDGLIFFVHDENTGKSCVTVIDAKTMSAPVAVVELPHRV 518
QY 581 PYGFHGTGFIAGDADLAKOV 599
Db 519 PYGFHALFVTEOLOEQTL 537
RESULT 10
US-09-758-269-33
; Sequence 33, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 33
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-33

Query Match 29.7%; Score 936; DB 10; Length 538;
Best Local Similarity 37.0%; Pred. No. 3.9e-77;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;
QY 64 KQSSNPAIVVVKP-AKESNTKOMNLFORAAALDAAGFLVSHKHLPLKTAQPSVQ 122
Db 4 KLDGSIITISVHPRKSGFSLDLRLVVKLM-----HDSPLH-----Y 47
QY 123 IAGNAPV-NEQVRRNLPLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAVK 181
Db 48 LSGNEAPIRDETPVKDLPVHGFLPECLNGEFVRVGNPKFDVAGVHNFDDGDMHIGVR 107
QY 182 FEHGSYACRFTQTNRFVQERQLGRPVFPAIGELHGHGTGIARLMLFYARAAAGIVDPA 241
Db 108 IKDGKATVYSRYVKTSLKQEEFPGAAPKMK-IGDLKGFGLLMVNVQQLFKLKILDNT 166
QY 242 HGTGVNAGVLVFNGLLMSDDLPOVOITPNDLKTGVREFDGOLESTMIAPKVD 301
Db 167 YNGNTANTALVYHGLKALLAQEADKPVVIXVLEDGLQTLIGIDYKRLHSTHAPKVD 226
QY 302 PESGEFLSYDVVSPYLKYFRFSPDGTSPDVEIQDQTMHDPATITENFVVVPDQ 361
Db 227 PVIGEMFTFGYS-HTPPLYLYRVLSKDGIMHDPVITISEPIMHDPATITETVAIENDLP 285
QY 362 VVFKLPMIRGGSPV-YDNKNVARGILDKYAEDSSNIKWIDAPDCFCFLNNAEPE 420
Db 286 MHERPKEMVKEKKMIYFDPDKARFGVLPRYAKDELMIRWELPNCFFIFHNANAWP--E 343
QY 421 TDVVVIGSCMTDPD-----SIFNESDENLKSVLSEIRLNKLTGSETRRPITISNDQVN 475
Db 344 EDEVVLTICLENLDLDMVSGKVKAEKLENFCNELYEMRNFNMKTSASQKLSASA----- 398
QY 476 LEAGMVNRNMLGRKTKFAYLALAPWPVKSGFAKVD-----LTTGEVKKHLY-- 522
Db 399 VDFPRINECYTGCKQRYVYVGTILDSIAKVTGIKFDLHAEATGKRMLEVGNGIKGIYDL 458

QY 523 GDNRYGGEPLFLPGEGGEDEGYILCFVHDEKTKWSELOIVNAVSLVE--ATVKLPSRV 580
Db 459 GGRYGSEAIYVPRETAEDDGLIFFVHDENTGKSCVTVIDAKTMSAPVAVVELPHRV 518
QY 581 PYGFHGTGFIAGDADLAKOV 599
Db 519 PYGFHALFVTEOLOEQTL 537
RESULT 11
US-08-976-063C-22
; Sequence 22, Application US/08976063C
; Publication No. US20020182697A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLI
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: HP VECTRA
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,063C
; FILING DATE: 21-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 196 49 655.1 (Germany)
; FILING DATE: 29-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-063C-22
Query Match 13.1%; Score 411.5; DB 1; Length 505;
Best Local Similarity 25.8%; Pred. No. 4.6e-29;
Matches 137; Conservative 79; Mismatches 192; Indels 123; Gaps 19;
QY 118 DPSVQIAGNEAPVNEQPVRRNLPLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMV 177
Db 8 DP--QLVGTLLPRTIEADLFDLEVDGEIPEKINSINGTFVRNTPEPQVTPQKEHTFIDGDMGMA 65
QY 178 HAYKFEHGSYACRFTQTNRFVQERQLGRPVFPAIGELHGHGTGIARLMLFYARAAAGI 237
Db 66 SAFHFEDGHVDFISRWVKTARFETARLARKSLFGMYRNPNYTDOTSVKGL----- 114
QY 238 VDBAHTGTVNAGVLVFNGLLMSDDLPOVOITPNDLKTGVREFDGOLES-TMIA 296
Db 115 -----DRTVANIISIHHGKVLAVKEDGLPY--ELDPR-TLETRGHFDYDQVTSQHTA 166
QY 297 HPKVPDPSGSEGLFALSYDVWSKPYLKYFRFSPDGTGKSPDV-----ETIQDQP- 342
Db 167 HPKYDPTGTGLL-----PFGSAKGEATFDMAYIYVDKHGKVTHTWPEQPY 213

Db 283 KEVSTKTYTDA--LVLXHHINAYEE--DGHVVEFDIVAYRDNLSYDMFYLLKKLKDREVN 337
QY 443 ENLKSVL-----SETRLM-----KTGSTRPPIISNQOQVNL 476
Db 338 NKLTISFTCKRFVVPLOYDKDAEYSGSLVKLPTSATAVKKEKDGSIYQOPELICE---GI 393
QY 477 EAGNVNMLGRKTKFAYLALAEWPKVSGFAKVDLTGTGEVKKHLYGDNRYGGEPFLPLPG 536
Db 394 ELPRVNDYNGKKYKYVAYEVQSPVPTKIAKLNVTQKEV-LHWGEDHCWSPSPFVPS 452
QY 537 -EGGEDEGVIL-CFVHDEKTKWSELOIVNAVSLEV--EATVKLPSPVPGFHGTF 588
Db 453 PDAREDEGVVLTCCVVVSEPNKAPFLILLDAKTEKELGRATVNVEMHL--DLHGMF 506

RESULT 14

US-10-053-192-5
; Sequence 5, Application US/10053192
; Publication No. US20030087336A1
; GENERAL INFORMATION:
; APPLICANT: BACHMANN, Heinrich
; APPLICANT: BRUGGER, Roland
; APPLICANT: FRIEDLEIN, Arno M
; APPLICANT: WIRTZ, Gabriele M
; APPLICANT: WOGGON, Wolf-Dietrich
; APPLICANT: WYSS, Adrian
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: BETA, BETA-CAROTENE 15, 15'-DIOXYGENASES, NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE
; FILE REFERENCE: B, B-CAROTENE 15, 15'-DIOXYGENASES, ...
; CURRENT APPLICATION NUMBER: US/10/053, 192
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 103382.0
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 529
; TYPE: PRT
; ORGANISM: BOVINE
US-10-053-192-5

Query Match 5.9%; Score 185.5; DB 9; Length 529;
Best Local Similarity 20.2%; Pred. No. 2.7e-08;
Matches 112; Conservative 91; Mismatches 209; Indels 143; Gaps 26;

QY 141 VVGKLPDSIKGVYVRNGANP-----LHEPVTGHHFFDGDGMVHAVKPEHGSASVACRFTQ 195
Db 11 VTGRIPLWLGSLLRCTGTGPLEVSGEPF--YHFDGQALLHKFDEKGEHVTYHPEIR 68
QY 196 TNRFVQERQLGRPVFKAIGELHCHTGIARMLFYARAAAGIVDPAHGTGVANAGLYFN 255
Db 69 TDAYVRAMTEKRIIVITE-----FGFTTCAFPDPCC--KNIFSRFFSYER 109
QY 256 G-----RLAMSED-----DLPYQVQILTPNGDLKTVGREFDGOLE-----STM 294
Db 110 GVEYTDNALNVIPVGDGEYACTETNITKINPE-TLEII-----FTKQVDLCNVSYNGA 164
QY 295 IAHPKVDPPES-----GELFALSVDVSKPKYLKVFRESPOGKTSPDVEIQID----- 340
Db 165 TAPHIENDGTGVNIGNCFGNKFSIANIVKIPLOADKEDPISKFSIENVQPCSDRF 224
QY 341 OPTMMHDFATENVVVDQOVFKLEPMIRGGSPVYDKNKVARF-----GI 388
Db 225 KPSYVHSGFLTPNIVVETPVKINLEKFLSSWS--LWGANVMDCFESFTNETMGVWLHI 282
QY 389 LDKYAEDESSNLKWDAPDCFCFHLNNAWEPEDEVVVIGSCMTPPDSINESD---ENL 445
Db 283 ADKARKKYLANKYKTSF--FNLFHINTYE--DNGFLIVLCCWKGFDEVINYFTLYLANL 339
QY 446 KSVLSEIRLNKLTGSTRRPTISNQOQVNLKAGMWRNM--LGRKTKFAYLA-----496

Db 340 RENWEEVK---KNARKAPQEPVRRVYVPLNIDKADTKGNLVTLPNTATATILCSDEFTTI 396
QY 497 LAEPWPKVSG-----FAKVDLT--TGEVKKHLYG----- 523
Db 397 WLEPEVLSGPRQAFEPFQINQYQYCGKPYTYAYGLNLHNFVDPDLCKLNKTKETWFTV 456
QY 524 --DNRYGGEPFLFP--GEGGEDEGYILCFVHDEKTKW--SELOIVNAVSL-----EVEAT 573
Db 457 WOEPDSTSPSEPIFVSHPDALDEDDGVVLSVVVSPGAGCKPAYILLILNAKDLSEVARAEFT 516
QY 574 VKLESRYVPYGFHGT 588
Db 517 VEI--NIPVTFHGLF 529

RESULT 15

US-09-759-508B-2
; Sequence 2, Application US/09759508B
; Publication No. US20020182599A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C.
; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
; FILE REFERENCE: 00786/381002
; CURRENT APPLICATION NUMBER: US/09/759, 508B
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 26926
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-508B-2

Query Match 3.4%; Score 107; DB 9; Length 26926;
Best Local Similarity 19.0%; Pred. No. 2.2e+02;
Matches 131; Conservative 97; Mismatches 205; Indels 256; Gaps 38;

QY 51 SSALHTPPAL--HPPKQSSNSPAI--VVKPKAKESNTKOMNL-----FORAAAAALDAAE 103
Db 15827 STYLNSEPTVAQYFVKVPGPGPGTPTVLTSSRDSMEVQWNEPISDGGSRVIGYHLERKERN 15886
QY 104 LVSEHEKLIH--PLPKTADPSVQIAGNFAPVNEQPVV-----RNLVPVGLPDSIKGVYV 154
Db 15887 SILWVKLNKTPITQTKTKTGL-----EGVEYEFVRSAAENIVGIGK--PSKVSECV 15937
QY 155 -RNGANPLHEPVTGHHFFEDGDGMV---HAVKFEHGSASVACRFTQTNRFVQERQLGRPV 210
Db 15938 ARDPCDPPGRP-----EALIVTRNSVTLOWKKPTYDGGSKITGYIVKEKKELPEGRW 15988
QY 211 PKA-----IGELHCHTGIARMLFYARAAAGIV--DPAHGTGVANAGLVYENGR 257
Db 15989 MKASFNTIIDTFEVTGLVEDHRYEFVFI---ARNAGVSESESTGA----- 16034
QY 258 LLAMSEDDLPYQVOITPENGDLKTVGREFDGOLESTMIHAPKVDPESEGFALSVDVSK 317
Db 16035 ITARDEVDP-----RISMDPKYKDTIVVH-----AGESFKVDADIYK 16073
QY 318 PY--LKVFRSPDGTCKSPDVEIQ-----LDQPTMMHDFATEN 353
Db 16074 PIPTQWIKGQDELNSFARLEIKSTDFATSLSVKDAVRVDSGNVYLAKKAVAGERSVTN 16133
QY 354 FVVV-----PDQOVVEK-----LPEMIRGGSPVYDKNKVARFGILOKYAEDSS 397
Db 16134 VKVLDREPPEGPVVISVTAECTLAWKPLQDGGSDII--NYIVE-----RRETS 16183
QY 398 NIKW--IDAP-----DCFCFHLNNAWE-----EPETDEVVWIGSCMTPPDS- 436
Db 16184 RLWMTVVANDVQTLSCVKTKLLEGNETFRIMAVNKYGVGEPLSEPVVAKNPFVVPDAP 16243
QY 437 -----IFNESDEN-----LKSVLSEIRLNKLT 458

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:15 ; Search time 21 Seconds
(without alignments)
2742.120 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150
Sequence: 1 MASTATAAAGSRWLGNGHFT.....VPGFCHGTFIGADDLAKQV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2289	72.7	604	2 T51936	probable 9-cis-epo
2	2280.5	72.4	605	2 T07123	nine-cis-epoxycaro
3	2170.5	68.9	657	2 E96812	protein F9f9.10 [
4	2044.5	64.9	589	2 A86425	probable 9-cis-epo
5	1991	63.2	583	2 T04531	nine-cis-epoxycaro
6	1930	61.3	604	2 T04351	viviparous-14 prot
7	968	30.7	595	2 T04438	hypothetical prote
8	938	29.8	538	2 T49193	neoxanthin cleavag
9	937	29.7	538	2 T51734	neoxanthin cleavag
10	745.5	23.7	475	2 AG1944	hypothetical prote
11	745	23.7	446	2 T17019	hypothetical prote
12	528.5	16.8	483	2 E87345	conserved hypothet
13	466	14.8	501	2 A70534	hypothetical prote
14	454.5	14.4	502	2 A70582	hypothetical prote
15	394	12.5	490	2 S76169	hypothetical prote
16	355.5	11.3	497	2 AE2341	lignosilbene-alph
17	352	11.2	484	2 JN0595	lignosilbene-alph
18	317	10.1	480	2 S76206	hypothetical prote
19	303.5	9.6	489	2 JCA324	lignosilbene-alph
20	253	8.0	616	2 T10688	hypothetical prote
21	244	7.7	618	2 C84885	hypothetical prote
22	242	7.7	472	2 AG2417	hypothetical prote
23	238.5	7.6	556	2 F88115	protein F53C3.12 [
24	231	7.3	483	2 D87290	conserved hypothet
25	215	6.8	533	2 A47143	retinal pigment m
26	126	4.0	790	2 T34293	hypothetical prote
27	116.5	3.7	1832	2 T31113	mucin-like glycop
28	114	3.6	870	2 S74291	hypothetical prote
29	113	3.6	466	1 IQEBV	replication initia

30 110.5 3.5 658 2 S60170 protein kinase Pak
31 108 3.4 437 2 G70019 conserved hypothet
32 107.5 3.4 658 2 T39500 serine/threonine-s
33 107 3.4 1461 2 T13157 mitotic checkpoint
34 107 3.4 26926 1 I38344 titin, cardiac mus
35 106.5 3.4 1299 2 T47182 hypothetical prote
36 106 3.4 1679 2 T15968 hypothetical prote
37 106 3.4 3562 2 A47171 chondroitin sulfat
38 105.5 3.3 2591 2 T30288 pristinamycin I sy
39 105 3.3 2504 1 A57788 enoyl-facyl-carrie
40 104.5 3.3 623 2 T52293 MYC-related DNA bi
41 104.5 3.3 1080 2 A35088 phycobllisome link
42 103.5 3.3 594 2 AG3616 hypothetical prote
43 103 3.3 775 2 B70449 phenylalanine-tRNA
44 103 3.3 4845 2 T31067 BIR repeat contain
45 102.5 3.3 531 2 AH3378 cell wall degradat

RESULT 1
T51936
probable 9-cis-epoxycarotenoid dioxygenase [imported] - potato
C:Species: Solanum tuberosum (potato)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T51936
R:Burridge, A.; Taylor, I.B.; Thompson, A.
A:Description: Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.
A:Reference number: Z25874
A:Accession: T51936
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-604 <BUR>
A:Cross-references: EMBL:AJ276244; PIDN:CAB76920.1
C:Genetics:
A:Gene: nced1

Query Match 72.7%; Score 2289; DB 2; Length 604;
Best Local Similarity 71.3%; Pred. No. 1.3e-165;
Matches 432; Conservative 66; Mismatches 90; Indels 18; Gaps 5;

QY 5 TATAAAGSRWLGNGHFTPTLSSQSSDLSYCS--SLPWASRVTRKLNVSALHTTTPALH 61
DB 3 TTTSHTAINTWI-----KPKLSMPSSKEFGFASNSILLKNQHRQSLNINSSLAQAPILH 57
QY 62 FPKQSSN-----SPAIVYKPKAKESNT-----KOMNLFQRAAAALDAAGFLVSHEKLHPL 113
DB 58 FPKQSSNYQTPKTTSTISHPKQENNNSSSSISKNLVQKAAAMALDAVEGALTKEHEHPL 117
QY 114 PKTADPSVOIAGFAFNPVEQPVRRNLPVYVKLPDSIKGVYVRNGANPLHEPVTGHHFFDG 173
DB 118 PKTADRVQLSGNFAPVNPENVCQSLPVTGKIPKCVYVRNGANPLFEPTAGRFFDG 177
QY 174 DGMVHAVKEHGSASACRFTQTNRFVQERQLGRPFVPEKAIAGELHGTGIARLMLFYARA 233
DB 178 DGMVHAVQVKNQSGASACRETETETEFVQEKALGRPFVPEKAIAGELHGTGIARLMLFYAR 237
QY 234 AGIVDPANGTGVANAGLVYFNGLRLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLEST 293
DB 238 LFLGLIHSRGTGVANAGLVYFNGLRLAMSEDDLPYHVKVTPTGDLKTEGRFDFDGOLEST 297
QY 294 MIAHPKVDPESGELFALSVDVNSKPKLYKFRFSPDGTSPDVEIQLDQTMHDFAITEN 353
DB 298 MIAHPKLDPSVSGELFALSVDVNIQPKLYKFRFSGNKEKSNVDVEIPDEPTMMHDFAITEX 357
QY 354 FVWVPDQVQVFKLPEMIRGSGSPVYDKNKVARGLDKYKAEDSSNKKWIDAPDCFCFLHW 413
DB 358 FVLIIPDQVQVFKSEMIIRGSGSPVYDKNKVSRLGILDKYKAGSDLKWKVEVPDCFCFLHW 417
QY 414 NAMEEPTEDVYVIGSCMTTPDPSIFNESDENLKSVLSEIRNLKGTGSTRRPTISNEQQ 473

ALIGNMENTS

Db 418 NAWPEETDELIVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGKSTRKAIENPDEQ 477
 QY 474 VNLEAGMVRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPLF 533
 Db 476 VNLEAGMVRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPLF 537
 QY 534 LPGE--GGEDEGYILCFVHDEKWKSELOQIVNAVSLVEATVKLPSRVPGFHTGTFGA 591
 Db 538 LRPDPNSKEEDGYILAFVHDEKWKSELOQIVNAVSLVEATVKLPSRVPGFHTGTFGA 597
 QY 592 DDLAKQ 597
 Db 598 NDLANQ 603

RESULT 2
 T07123
 nine-cis-epoxycarotenoid dioxygenase - tomato
 N:Alternate names: probable neoxanthin cleavage enzyme
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
 C:Accession: T07123
 R:Burridge, A.
 submitted to the EMBL Data Library, January 1998
 A:Reference number: Z15934
 A:Accession: T07123
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-605 <BUR>
 A:Cross-references: EMBL:Z97215; PIDN:CAB10168.1

Query Match 72.4%; Score 2280.5; DB 2; Length 605;
 Best Local Similarity 71.7%; Pred. No. 6e-165;
 Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

QY 5 TATAAAGVSGWLGNGHTQPLSSQSSDLSVCS---SLPMASRVTRKLVSSALHTP2ALH 61
 Db 3 TTTSHATNWI-----KTLKMPSSKEFGFASNSISLLKNQHNROSLNINSLSQA2PILH 57

QY 62 FPKQSN--SPA--IVVKPKAKESN-----TKQNNIFQRAAAALDAAGFLVSHKELHP 112
 Db 58 FPKQSNQYQTKNNITISHPKQENNNSSSTSKNVLQKAAALDAVSA2LTKHELEHP 117

QY 113 LPKTADPSVOIAGNPAPVNEQVRNLPVVKLPISIKGVYVNGANLPHBPVGTGHFFED 172
 Db 118 LPKTADPRVOISGNFAPVNEQVRNLPVVKLPISIKGVYVNGANLPHBPVGTGHFFED 177

QY 173 GDMVHAVKFEHGSASACRFTOTNRFVQEROLGRPVFPKALGELHGTGIALMLFYAR 232
 Db 178 GDMVHAVOFKNGSASACRFTETERLVQEKALGRPVFPKALGELHGTGIALMLFYAR 237

QY 233 AAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDLPYQVOITPBGDLKTVGRFDDGQLES 292
 Db 238 GLFGLVDHSGKGVANAGLVYFNGRLLAMSEDDLPYQVOITPBGDLKTVGRFDDGQLES 297

QY 293 TMIAHPKVPDSGELFALSVDVSKPYLYKFRSPDGTSPDVEIQDLOPTMMHDFATE 352
 Db 298 TMIAHPKLPDPSGELFALSVDVSKPYLYKFRSPDGTSPDVEIQDLOPTMMHDFATE 357

QY 353 NFVVVDDQVQVFKLPMIRGGSPVYVYDKNKVARFGLDKYAEADSSNIKWIDAPDFCFLH 412
 Db 358 NFVVIPDQVQVFKMSEIRMGSPVYVYDKNKVARFGLDKYAEADSSNIKWIDAPDFCFLH 417

QY 413 NNAWEPEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGKSTRKAIENPDEQ 477
 Db 418 NNAWEPEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGKSTRKAIENPDEQ 477

QY 473 QVNLEAGMVRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPL 532
 Db 478 QVNLEAGMVRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPL 537

QY 533 FLPGE--GGEDEGYILCFVHDEKWKSELOQIVNAVSLVEATVKLPSRVPGFHTGTFGA 590
 Db 538 LRPDPNSKEEDGYILAFVHDEKWKSELOQIVNAVSLVEATVKLPSRVPGFHTGTFGA 597

Db 538 FLPRDPNSKEEDGYILAFVHDEKWKSELOQIVNAVSLVEATVKLPSRVPGFHTGTFIN 597
 QY 591 ADDLAKQ 597
 Db 598 ANDLANQ 604

RESULT 3
 E96812
 protein F3f9.10 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E96812
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 ansen, C.W.; Hughes, B.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mailli, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: E96812
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-657 <STO>
 A:Cross-references: GB:AE005173; NID:g8052533; PIDN:AAF71797.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F3f9.10
 A:Map position: 1

Query Match 68.9%; Score 2170.5; DB 2; Length 657;
 Best Local Similarity 67.7%; Pred. No. 1.5e-156;
 Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;

QY 1 MASFTATAAASVRGWL-----GGNHTQPLSSQSSDLSVCSLSP-----MASRVT 45
 Db 51 MASTILLPSTSTQFLDRTSTSSSRPKLQS-----LSFSTLRNKKLVVPCYSSSVN 105

QY 46 RKLNVSSALHT----PPALHFFKQSSNSPAIVVKKAKESNTKOMLFORAAAAALDAE 101
 Db 106 KKSSVSSLSQSPTEPPSW---KKLCNDVTNLI-PKTTNQNK-LNPVQRTAMVLDVE 160

QY 102 GLVSHK-KLHLPKPTADPSVOIAGNFAVNEQVRNLPVVKLPISIKGVYVNGANP 160
 Db 161 NAMI SHERRRHPKPTADPAVQIAGNFFVPEPKPVVHNLPTVGTVECTQGVYVNGANP 220

QY 161 LHEPVTGTHFFDGDGMVHAVKEHGSASVACRFTOTNRFVQEROLGRPVFPKALGELHGH 220
 Db 221 LHKPVSGHHLFDGDMVHAVRFDNGSVSVACRFTETNRLVQERCGRPVFPKALGELHGH 280

QY 221 TGIARLMIFYARAAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDLPYQVOITPBGDLK 280
 Db 281 LGIAKLMFLNTRGLEFLVDPTGGLGVANAGLVYFNGHLLAMSEDDLPYHVKVTQDGLT 340

QY 281 VGRFDDGQLESTMIAPKVPDSGELFALSVDVSKPYLYKFRSPDGTSPDVEIQD 340
 Db 341 SGRYDFDQQLKSTMIAPKVPDSGELFALSVDVSKPYLYKFRSPDGTSPDVEIQD 400

QY 341 OPTMMHDFAITENFVVDPQVQVFKLPMIRGGSPVYVYDKNKVARFGLDKYAEADSSNIK 400
 Db 401 OPTMHDFAITENFVVIPDQVQVFKLPMIRGGSPVYVYDKNKVARFGLDKYAEADSSNIK 460

QY 401 WIDAPDCFCFHLNNAWEPEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGTE 460
 Db 461 WIEVPDCCFHLNNAWEPEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGTE 520

QY 461 STRPRPISNEQVQVFKLPMIRGGSPVYVYDKNKVARFGLDKYAEADSSNIK 520
 Db 521 STRPRPISNEQVQVFKLPMIRGGSPVYVYDKNKVARFGLDKYAEADSSNIK 577

QY 521 LYGNRYGGEPLFLPGEGEEDGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSRV 580
 Db 578 IYGEKYGGEPLFLPSGDEGEGYIIVFVHDEKVKSELOIVNAVNNKLEATVILPSRV 637
 QY 581 PYGFHGTIFGADDLAKQVY 599
 Db 638 PYGFHGTIFISKEDLSKQAL 656

RESULT 4

A86425
 probable 9-cis-epoxycarotenoid dioxygenase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: A86425
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: A86425
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-589 <SFO>
 A:Cross-references: GB:AE005172; NID:g11094779; PIDN:AAG29711.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 64.9%; Score 2044.5; DB 2; Length 589;
 Best Local Similarity 68.1%; Pred. No. 4.7e-147;
 Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps 4;

QY 27 SQSSDLSYCSLPMASRVTRKLVSSALHTPPALHFFKQSSNSPAIVVVKPAKESNTKQM 86
 Db 27 SPSSVSFTWKPR---RRKLSANSVSDTPNLLNFNFYSPNII-----PEKQTSRW 76
 QY 87 NLFQRAAAALDAEGLVSHKHLPLKPTADFSVQIAGNFAPVNEQPVRRNLPVVGKLP 146
 Db 77 NPLQRAAALDAEATALLRRERSKPLKPTVDPHQISGNYAPVPEQSVKSSLSVDGKIP 136
 QY 147 DSIKGVYVRNGANPLHPTGHHFFDGDGMVHAVKFEHGSASACRTQTNRFVQERQLG 206
 Db 137 DCDIGVYLRNGANPLFEPVSGRHLDGDDGMVHAVKIITNGDASYSCRFTETRLVQERQLG 196
 QY 207 RPVFPAKAGELHGHGTIARLMFLYARAAGIVDPAHGTGVANAGLVYFNGRLAMSDDL 266
 Db 197 SPIFPAKAGELHGHGTIARLMFLYARGLFGLLNKNGTGVANAGLVYFHRLLANSEDDL 256
 QY 267 PYQVQITPNCGLKTVGRDFDGOLESTMIAPKVPDPSEGLFALSVDVSKPKLYKFRFS 326
 Db 257 PYQVRYTDNGDLETIGRDFDGOLESTMIAPKIDPVTKELFALSVDVKKPKLYKFKFS 316
 QY 327 PDGKTPDVEIQDQPTMMDHFAITENFVVVDPQVVEKLPPEMIRGSPVYVQKNAVRF 386
 Db 317 PEGEKSPDVEIQPTMMDHFAITENFVVVDPQVVFKLSDPFLGKSPVKYDGEKISR 376
 QY 387 GILDKYAEPSNTKWIADPCFCFHLNNAWEEPETDEVVVGSCMPPPOSIFNESDENLK 446
 Db 377 GILPRNAKDAEMVVESEPTFCFHLNNAWESPETDEVVVGSCMPPADSIFFNECDQLN 436
 QY 447 SVLSEIRLNKGTSESTRPIISNEDQVNLNLEAGVNRNMLGRKTKFAYLALAPWPKVSG 506
 Db 437 SVLSEIRLNKGTSESTRPIIPG-SVQMNLEAGVNRNMLGRKTRVAYLALAPWPKVSG 495
 QY 507 EAKVDLTTCGEVKKHLYGDNRYGGEPLFLPG--EGGEDEGYILCFVHDEKTKWSELOIVN: 564
 Db 496 FAKVDLTGEVKNHFGYGGKYGGEPPFLPRGLSDGDDGYIMSFVHDESESELIHVN 555

QY 565 AVSLEVEATVKLPSRVYPYGFHGTIFGADDLAKQ 597
 Db 556 AVTLELATVKLPSRVYPYGFHGTIFVNSADMLNQ 588

RESULT 5

T04531
 nine-cis-epoxycarotenoid dioxygenase homolog F28J12.10 - Arabidopsis thaliana
 N:Alternate names: hypothetical protein T9A21.200
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C:Accession: T04531; T04937
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancr
 submitted to the Protein Sequence Database, February 1998

A:Reference number: Z15377
 A:Accession: T04531
 A:Molecule type: DNA
 A:Residues: 1-583 <BEV>
 A:Cross-references: EMBL:AL021710
 A:Experimental source: cultivar Columbia; BAC clone F28J12
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: Z15390
 A:Accession: T04937
 A:Molecule type: DNA
 A:Residues: 1-377 <BEW>
 A:Cross-references: EMBL:AL021713
 A:Experimental source: cultivar Columbia; BAC clone T9A21
 C:Genetics:
 A:Map position: 4
 A:Note: F28J12.10; T9A21.200

Query Match 63.2%; Score 1991; DB 2; Length 583;
 Best Local Similarity 64.5%; Pred. No. 5.3e-143;
 Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;

QY 16 GGNHTQPLSSQSSDLSY--CSSLPMASRVTRKLVSSALHTPPALHFFKQSSNSPAIV 73
 Db 11 GGKWTW---QAQIDLGFRPIKQPKVICTQVIDVTE-LTKKKQLFTPTTATTP--- 62
 QY 74 VKPAKESNTKOMNLFQRAAALDAEGLVSHKHLPLKPTADFSVQIAGNFAPVNEQ 133
 Db 63 -----QHNELRLNIFQRAAALDAEALALISHEQDSPLPKTADPRVQIAGNTPV 116
 QY 134 PVRNLPVVGKLPDSIKGVYVRNGANPLHPTGHHFFDGDGMVHAVKFEHGSASACRF 193
 Db 117 SVRNLTVEGTIPDCIDSVIRNGANPMFEPTAGHLEFDGDMVHAVKIITNGSASACRF 176
 QY 194 TQTNRFVQERQLGPPVFPKAGELHGHGTIARLMFLYARAAGIVDPAHGTGVANAGLVY 253
 Db 177 TKTERLVQERKLRGPVFPKAGELHGHGTIARLMFLYARGLCGLINNQNGVGVANAGLVY 236
 QY 254 FNGRLAMSDDLPIYQVQITPNGDLKTVGRDFDGOLESTMIAPKVPDPSGELFALSVD 313
 Db 237 FNNRLAMSDDLPIYQLKIITQDQTVGRDFDGOQKRSAMIAHPKLDPVTKELHALSYD 296
 QY 314 VSKPKLYKFRFSPDGTSPDVEIQDQPTMMDHFAITENFVVVDPQVVEKLPPEMIRGG 373
 Db 297 VVKPKLYKFRFSPDGVKSPLEIPLTPTMIDHFAITENFVVVDPQVVEKLPPEMIRGG 356
 QY 374 SPVYDKNKVARFGLDKYAEPSNTKWIADPCFCFHLNNAWEEPETDEVVVGSCMTP 433
 Db 357 SPVYDEKVKYSLRGIMPKDATEASQIIVWNSPETECFHLNNAWESPETEIVVIGSCMSP 416
 QY 434 PDSIFNESDENLKSVLSEIRLNKGTSESTRPIISNEDQVNLNLEAGVNRNMLGRKTKFA 493
 Db 417 AUSIFNERDESUKSVLSEIRLNKGTSESTRPIISNEDQVNLNLEAGVNRNMLGRKTKFA 474
 QY 494 YLALAEFWPKVSGFAKVDLTTCGEVKKHLYGDNRYGGEPLFLPGEGG---EEDGEGYILCF 549
 Db 475 FLAIAYFWPKVSGFAKVDLTTCGEMKKYITGEGYKGGEPFPLPNSNGENEGDDGYIFCH 534

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-538 <RIE>
A:Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.150
A:Experimental source: cultivar Columbia; BAC clone MAA21
C:Genetics:
A:Gene: ATSP:MAA21.150
A:Map position: 3
A:Introns: 72/3; 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 377/2; 411/2; 49

Query Match 29.8%; Score 938; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 4e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 49; Gaps 13;

QY 64 KSSNSPAIVVVKPK-AKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLPKTADPSVQ 122
DB 4 KLSGSSIIIVHPRSPKSGFSSKLLDLLELVVKLM-----HDASLPLH-----Y 47
QY 123 IAGNFAPV-NEQPVARNLPVVGKLPDSIKGVYVRNGANPLHEPVTHGHFFDGDGMVHAVK 181
DB 48 LSGNFAPIRDETTPPVKDLPHVGHFLPECLNGEFVRGPNPKFDVAGYHWFDDGDMHIGVR 107
QY 182 FEHGSASACRFTQTNRFVQERQGRPVFPKATIGELHGHGTARLMLFYARAAAGIVDPA 241
DB 108 IKDKATIVSVRYVKTSLRKQEEFFGAAKFMK-IGDLKGFGLLMVNVQQLRTKLKILNT 166
QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNDGLKTVGRDFDGOLESTMIHAPKVD 301
DB 167 YGNTANTALVYHGGKLLALQADKPYVVKVLEDGLQTLIGIYDKRLTHSFTAHKVD 226
QY 302 PESGELFALSVDVSKPYLKYFRSPDGTSPDVEIQLDPTMMHDEFAITENFVVVPDQ 361
DB 227 PVTGEMTFEGYS-HTPPYLYTVRISKDGMHDPVITISEPIMHDFAITETIYAFMDLP 285
QY 362 VVFKLPEMIRGSPV-YDKNKVARFGLDKYAEDESSNIKWIDAPDCFCFHLNNAWEEPE 420
DB 286 MHFPRKEMVKEKMIYSPDPTKKARFGLVPRYAKDELIRWELNCFIFHNANAE--E 343
QY 421 TDEVVVIGSCMTPPD-----SIFNESDENLKSVLSEILNLTGSTRPIISNEDQOVN 475
DB 344 EDEVVLTICRLNFDLDMVSGVKVEKLENGELNEMFNKMGTSASOKLSASA----- 398
QY 476 LEAGWVRNMLGRKTKFAYLALAEWPVKVSGFAKVD-----LTTGEVKKHLY-- 522
DB 399 VDFPRINECTYTKKQRYVYGTILDSIAKVTGIIFDLHAEAEKGRMLEVGNKIGIYDL 458
QY 523 GDNRYGGEPLFLPGEGEDEGYILCFVHDEKTKWSELQIVNAVSLVE--ATVKLPSRV 580
DB 459 GEGRYGSEAIYVPREAEEDDGYLIFFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518
QY 581 PYGFHGTFIGADDLAKQVV 599
DB 519 PYGFHALFVTEEQLEQOTL 537

RESULT 9
T51734
neoxanthin cleavage enzyme [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51734
R:Neill, S.J.; Burnett, E.C.; Desikan, R.; Hancock, J.T.
J. Exp. Bot. 49, 1893-1894, 1998
A:Title: Cloning of a wilt-responsive cDNA from an Arabidopsis thaliana suspension culture
A:Reference number: Z24454
C:Accession: T51734
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-538 <NEI>
A:Cross-references: EMBL:AJ005813; PIDN:CAA06712.1
A:Experimental source: Cultivar Landsberg erecta
C:Genetics:
A:Gene: ncl

Query Match 29.7%; Score 937; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 4.8e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KSSNSPAIVVVKPK-AKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLPKTADPSVQ 122
DB 4 KLSGSSIIIVHPRSPKSGFSSKLLDLLELVVKLM-----HDASLPLH-----Y 47
QY 123 IAGNFAPV-NEQPVARNLPVVGKLPDSIKGVYVRNGANPLHEPVTHGHFFDGDGMVHAVK 181
DB 48 LSGNFAPIRDETTPPVKDLPHVGHFLPECLNGEFVRGPNPKFDVAGYHWFDDGDMHIGVR 107
QY 182 FEHGSASACRFTQTNRFVQERQGRPVFPKATIGELHGHGTARLMLFYARAAAGIVDPA 241
DB 108 IKDKATIVSVRYVKTSLRKQEEFFGAAKFMK-IGDLKGFGLLMVNIQQLRTKLKILNT 166
QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNDGLKTVGRDFDGOLESTMIHAPKVD 301
DB 167 YGNTANTALVYHGGKLLALQADKPYVVKVLEDGLQTLIGIYDKRLTHSFTAHKVD 226
QY 302 PESGELFALSVDVSKPYLKYFRSPDGTSPDVEIQLDPTMMHDEFAITENFVVVPDQ 361
DB 227 PVTGEMTFEGYS-HTPPYLYTVRISKDGMHDPVITISEPIMHDFAITETIYAFMDLP 285
QY 362 VVFKLPEMIRGSPV-YDKNKVARFGLDKYAEDESSNIKWIDAPDCFCFHLNNAWEEPE 420
DB 286 MHFPRKEMVKEKMIYSPDPTKKARFGLVPRYAKDELIRWELNCFIFHNANAE--E 343
QY 421 TDEVVVIGSCMTPPD-----SIFNESDENLKSVLSEILNLTGSTRPIISNEDQOVN 475
DB 344 EDEVVLTICRLNFDLDMVSGVKVEKLENGELNEMFNKMGTSASOKLSASA----- 398
QY 476 LEAGWVRNMLGRKTKFAYLALAEWPVKVSGFAKVD-----LTTGEVKKHLY-- 522
DB 399 VDFPRINECTYTKKQRYVYGTILDSIAKVTGIIFDLHAEAEKGRMLEVGNKIGIYDL 458
QY 523 GDNRYGGEPLFLPGEGEDEGYILCFVHDEKTKWSELQIVNAVSLVE--ATVKLPSRV 580
DB 459 GEGRYGSEAIYVPREAEEDDGYLIFFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518
QY 581 PYGFHGTFIGADDLAKQVV 599
DB 519 PYGFHALFVTEEQLEQOTL 537

RESULT 10
AG1944
hypothetical protein all1106 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AG1944
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri:
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabat
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1944
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-475 <KUR>
A:Cross-references: GB:HA000019; PIDN:BA073063.1; PID:gl17130452; GSPDB:GN00179
A:Experimental source: Strain PCC 7120
C:Genetics:
A:Gene: all1106

Query Match 23.7%; Score 745.5; DB 2; Length 475;
Best Local Similarity 34.5%; Pred. No. 1.3e-48;
Matches 169; Conservative 91; Mismatches 171; Indels 59; Gaps 15;

QY 123 IAGNFAPVNEQPVARNLPVVGKLPDSIKGVYVRNGANPLHEPVTHGHFFDGDGMVHAVK 182
DB 48 LSGNFAPIRDETTPPVKDLPHVGHFLPECLNGEFVRGPNPKFDVAGYHWFDDGDMHIGVR 107

Db 22 LEGNAPVHEETDTTLAKVIGELPELSSMEVRNGPNQWPTIGQYHWFEDGDMLGVR 81
QY 183 EHGSSASYACRTQTNRFRVQEROLGRPVFPKALGELHGTGTLARMLFYARAAAGIVD--- 239
Db 82 SNGKATYRNRYVTRQWQIEHAGQAIW-----IGLMEPPQ 117
QY 240 ---PAHGTGVANAGLVYFNGRLAMSEDDLPVOVITPBGDLKTVGDFDQGFESTMIA 296
Db 118 TELPSRNTG--NTALIWHAGQALLAWEGAPYAIQVP---DLASIGEYTYNNQLSSAF 172
QY 297 HPKVDPESEGFALSVDYVSVKPYLKYFRSPDGTSPDVEIQDQOPTMMHDAFAITENFV 356
Db 173 HPKVDPEVTEGMMFFOYS--FAPPYLHYVSVATGELVTRVPIDLPNGVMHDAFAITANVTI 231
QY 357 VPDQVQVFKLPMIRGSPVYDKNKVARFGLDXYAEDSSNIKWIDAPDPCFHLWNAW 416
Db 232 FMDLPLTFVSERMQRGPEMLFESDRPGRGILPLRHG--DNSQIRWFEPAPSCVFTHLNAY 290
QY 417 EPEPTEVWVIG-----SCMTPPDSIFNESDENLK--SVLSEIRLNKLTGSTRPRIISN 469
Db 291 E--DKDEVVLFACPMRSTIVLASPPS---QTDPEADIPRLHWRHFKLTGKVHEEML--- 342
QY 470 EDQVNLKAGVYRNMLGRKTKFAYLA--LAE--PWKVSFGFAKVDLTGTGVKKHLYGDNRY 527
Db 343 --DDVASEFPRINENELGTOPTQYTSRLAKGSIFLFEGLIKYDLSNAKSNQYEGCGRY 400
QY 528 GCEPLFLPGEGE--EDEGYILCFVHDEKTKWSELOLVNAVSLEVE--ATVKLPSRVPGF 584
Db 401 GSEAVFVRPGATVEDDGLWLIYVYDTEGSESELVYVNAQDINSEPIARVLLPQRVPYGF 460
QY 585 HGCTFGADDL 594
Db 461 HGIWVTEQL 470

RESULT 11
T17019
Query Match 11
Best Local Similarity 39.8%; Score 745; DB 2; Length 446;
Matches 150; Conservative 76; Mismatches 129; Indels 22; Gaps 9;
R.Watillon, B.; Kettmann, R.; Arredouani, A.; Hequet, J.; Boxus, P.; Burny, A.
Plant Mol. Biol. 36, 909-915, 1998
A:Title: Apple messenger RNAs related to bacterial lignostilbene dioxygenase and plant
A:Reference number: Z18655; MUID:98179104; PMID:9520281
A:Accession: T17019
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-446 <WAT>
A:Cross-references: EMBL:Z93765; NID:g2924324; PIDN:CAB07784.1; PID:g2924325
A:Experimental source: cv. McIntosh, strain Wj/cik

Query Match 23.7%; Score 745; DB 2; Length 446;
Best Local Similarity 39.8%; Score 745; DB 2; Length 446;
Matches 150; Conservative 76; Mismatches 129; Indels 22; Gaps 9;
QY 231 ARAAGIVDPHAGTGVANAGLVYFNGRLAMSEDDLPYQVITPBGDLKTVGDFDQGL 290
Db 76 ARVLTGOYNPANGIGLANTSLAFFGDRLYALGESDLPYSLRLTNSGDIETLGRHDFDGL 135
QY 291 ESTMIHAKVPDPSEGLFALSVDVSKPYLYFRFSPDGTSPDVEI--QLDQPTMMHDA 349
Db 136 SMNTAHKPIDPTGEAFAPRYGFI--RPFLTYFRFSDNGYKQDPVIFSMVTTFELHDA 194
QY 350 ITNEFVVPDQVQVFKLPEMI--RGGSPVYDKNKVARFGLDKYAESNNIKWIDAPDCF 408
Db 195 ITRKHAIFAQIGLNLIDMTKRAPTFGDLPKVPRIQVPIYAKDESMRWFEPGFN 254
QY 409 CFHLWNAWEPEPTEVWVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRPRIIS 468
Db 255 GVHATNWD--EDDAIVWAPNVLSEAHVLERVD--LVHCLVEKVRIDLKGTIVTRQPIST 311
QY 469 NEDQVNLKAGVYRNMLGRKTKFAYLALAEPPKVSFGFAKVDLTGGE-----VKKHLYG 523

Db 312 R-----NLDFAVINFAYLGRKNKYVYAAEGDPMKISGVWKLDSNVHEKCEIVASRMFG 366
QY 524 DNRVGCEPLFLPGE-----GGEDEGYILCFVHDEKTKWSELOLVNAVS--LEVEATVKLP 577
Db 367 PCGYGCEFFVAREPENPEADENGLFVSYVHDEKAGESRFLVMDAKSPQLDIVAAVRMP 426
QY 578 SRVPYGFHGTIFGADDL 594
Db 427 RRVPGFHLFVRESDL 443

RESULT 12
E87345
Query Match 16.8%; Score 528.5; DB 2; Length 483;
Best Local Similarity 30.9%; Pred. No. 4.1e-32;
Matches 160; Conservative 89; Mismatches 203; Indels 65; Gaps 22;
conserved hypothetical protein CC0776 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87345
R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:AE005673; NID:gl3422015; PIDN:AAK22761.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0776

QY 108 EKLHPLPKTADPS--VOIAGNAPVNEQFVRNLPV--CKLPDSIKGVVVRNGANPLHEP 164
Db 4 ERLPPVRTSLGPTNHPYMTGPTPQHEEVNAMDLVEGAIFADLDGVYLRNTEPVHDP 63
QY 165 VTGHHFDFDGDGMVHAVKFEHGSASYACRTQTNRFRVQEROLGRPVFPKALGELHGTGIA 224
Db 64 IGRYHFFDSDGMHLQIEFKGGAATYRNRRVTRCFEAEQFVNEGLW---GGLMDGPGVS 119
QY 225 RMLFYARAAAGIVDPAHGT--GVANAGLVYFNGRLAMSEDDLPYQ-----VOITPBGDL 278
Db 120 K-----RPGFG---AHGALKDSASTDIVHNGEAIATF-----YQCGEAYRLDPL--TL 163
QY 279 KTVGRFEDDQLESTMIAPHKVPDPESEGLFALSVDVSK--PYLYKFRFSPDGTSPDVE 336
Db 164 ENLGVASW--APLEG--VSAHPKVDTEATGELMFFNY---SKAWPYMHYGVVPGDKRKYQG 218
QY 337 IQLDQPTMMHDAFAITENFVVPDQVQVFKLPEMIRGSPVYDKNKVARFGLDKYAES 396
Db 219 VPLPGRLPHDMAFSSKYAILNDLPVFWQELMAROIHAVRLHKGIPSRFALVPR---EG 275
QY 397 SNTKWIDAPDCFHLWNAWEPEPTEVWVIGSCMTPP-----DSIFNESDEN 444
Db 276 GEPRWEAEPTYYLHNLNAYE--DGDEWLDGYFQEKPIRPLEGAPDGHGLMAYLDEH 333
QY 445 LKSVLSEI---RLNLTGSTRPRIISNEDQVNLKAGVYRNMLGRKTKFAYLALAE-- 500
Db 334 --SFLPKLHWRNLTGTEKHL-----DDRV--LFGFMFNQKYAKPRYAYSTTAKPG 386
QY 501 WPKVSGFAKVDLTGTGEVKKHLYGDNRYGGEPLFLPGE--EEDGYILCFVHDEKTKWSE 559
Db 387 WFLNFGVFKHDLLETGESWSIALPEGYASAPAPKVGAVDEDDGYLVSFIIIDENKASE 446
QY 560 LQIVNAVSLEVEATVKLPSPRVPGFHTFGADDLAK 596
Db 447 CLIVDAKRFVWCRIALPHKLSSGTHVTWAGREMLTK 483

[illegible]

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, Jun

Query Match 12.5%; Score 394; DB 2; Length 490;
Best Local Similarity 25.7%; Pred. No. 6.8e-22;
Matches 138; Conservative 93; Mismatches 190; Indels 116; Gaps 27;

QY 114 PXTADPSVQIAGNEAP-----VNEQPVRENL--PVVGKLPDSIKGYVRNCA----- 158
Db 5 PTTSEPSQR---SYSPQDWLRGVSQSQFQWVDEVSIPPDQGLTYRNGPGLLEIG 61

QY 159 -NPLHEPVTHHFFDGDGMVHAVKEP-HGSASYACFTQTNRFVQERQGRPVFPKAIKE 216
Db 62 DRPLKHP-----FDGDMVTAFKPGDGRVHFQSKFVRTQGYVEQKAGKMYRGVFGS 115

QY 217 LHGHTGIARLMLEYARAAAGIVDPAGH-----TCVANAGLVYFNRRLLAMSEDDL 266
Db 116 -----OPAGGWLKTTFDLRLKNIANITNYWGDRLIALWEGGQ 153

QY 267 PYQVOITPNCGLKTVGRDFDGOLE--STMIAPKVDPE-----GELFALSVDVVS--KP 318
Db 154 PHRELE--PSNLATIGLDLGLAEGQPLSAHPRIDPASTFDGGQPCYVTFSIKSSLSS 210

QY 319 YLKYFPSPDGKYSQDVEIQLDQPT-----MMHDFAITENFVVVPDQOVVFK-LPEN- 369
Db 211 TLTLLDLPQK-----LLRQKTEFPFGAFIHDFAITPHYAIFLQNNVTINGLPYLF 263

QY 370 -JRG-GSPVVDKKNVAREGILDKYAEDSSNWKWIDAPCFCHLWNAEPEETDEVVI 427
Db 264 GLRGAGECVQFHPDKPAQIILVPR---DGEIKRIPVQAGVFHHANAFE--ENKIIID 318

QY 428 GSCMTPPDSIFNESDENLKSIVLSEIKNLKTGESTR--RPIISNEDQOVNL---EAGM 480
Db 319 SICYNLSLPQV--DTGDGFRSTNFD---NLDPGQLWRFTIDPAAATVEKQOLMYSRCCEFPV 373

QY 481 VNRRNMLGRKTKFAYLALAE-----PWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPL 534
Db 374 VHPQQVGRPYRYVYMGAAHSTGNAP---LQAILKVDLESGETTLRSFAPHGFAGEPIEV 430

QY 535 PGEKG-EEDEGVLCFVHDEKWKSELOIVNA--VSLEVEATVKLPSRVPYGFHGT 588
Db 431 PREGGVAEDDGLWLLCLIIYKADLRSELVILDAODITAPAIATILKHKHIPPYPLHGSW 487

Search completed: May 19, 2003, 15:49:09
Job time : 25 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:14 ; Search time 27 Seconds
(without alignments)
920.160 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAAVSGRWLGNGH.....VPYGRHFGAGDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	120	3.8	616	1 NXFL_COTJA	P58797 coturnix co
2	114	3.6	870	1 YCS0_YEAST	P25623 saccharomyc
3	113	3.6	466	1 DNAA_PROMI	P22837 proteus mir
4	107.5	3.4	658	1 PAK1_SCHPO	P50527 schizosacch
5	106	3.4	3562	1 PGCV_CHICK	P90953 gallus gall
6	105	3.3	783	1 CAD5_MOUSE	P55284 mus musculu
7	105	3.3	2504	1 FAS_HUMAN	P49327 homo sapien
8	104.5	3.3	1079	1 AFCE_FREDI	P16566 fremyella d
9	103	3.3	775	1 SYFB_AQUAE	C67620 aquifex aeo
10	102	3.2	2493	1 CYAA_USTNA	P49606 ustilago ma
11	101.5	3.2	520	1 GLGS_BRANA	Q9m462 brassica na
12	101.5	3.2	565	1 SCRL_SCHPO	O14335 schizosacch
13	101	3.2	512	1 GLGT_VICFA	P52417 vicia faba
14	100	3.2	3770	1 ACVS_EMENI	P27742 emericella
15	99.5	3.2	383	1 PHVC_BACSU	Q31097 bacillus su
16	99.5	3.2	1256	1 ATLL_STAUI	P52081 staphylococ
17	98	3.1	741	1 MASZ_MYCTU	Q50596 mycobacteri
18	98	3.1	1788	1 YP72_CAEEL	Q05221 caenorhabdi
19	97.5	3.1	518	1 CP74_ARATH	P96242 arabidopsis
20	97.5	3.1	691	1 SLP1_YEAST	P20795 saccharomyc
21	97	3.1	385	1 OXDC_BACSU	O34714 bacillus su
22	97	3.1	636	1 KDPG_SCHPO	O14019 schizosacch
23	96.5	3.1	426	1 UCR2_SCHPO	P78761 schizosacch
24	96.5	3.1	503	1 DUTA_BACSU	P39581 bacillus su
25	96.5	3.1	862	1 SMP2_YEAST	P32567 saccharomyc
26	96.5	3.1	3587	1 TYCB_BACBR	O30408 b tyrocidin
27	96	3.0	462	1 DNAA_YERPE	Q829u7 yersinia pe
28	95.5	3.0	865	1 NRFA_PENUR	Q92269 penicillium
29	95.5	3.0	918	1 DNLL_RAT	Q9jhy8 rattus norv
30	95.5	3.0	1144	1 DP3A_NEIMB	Q9jx22 neisseria m
31	94.5	3.0	399	1 EX7L_CLOAB	Q97hd0 clostridium
32	94.5	3.0	590	1 HMAA_DROME	P29555 drosophila
33	94.5	3.0	638	1 GHR_RAT	P16310 rattus norv

34 94.5 3.0 778 1 RHP9_SCHPO
35 94.5 3.0 953 1 LKAB_PASHA
36 94.5 3.0 1046 1 RPOC_WEIHE
37 94 3.0 532 1 P72_MYCWY
38 94 3.0 919 1 X893_HUMAN
39 93.5 3.0 872 1 SYA_LACIA
40 93.5 3.0 1275 1 GNRP_HUMAN
41 93 3.0 3119 1 CALC_MOUSE
42 92.5 2.9 354 1 OTX1_HUMAN
43 92.5 2.9 746 1 PCAP_HUMAN
44 92.5 2.9 955 1 LKAA_PASHA
45 92.5 2.9 1057 1 SPSL_CITUN

ALIGNMENTS

RESULT 1

ID NXFL_COTJA STANDARD; PRT; 616 AA.
 AC P58797;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated protein) (mRNA export factor TAP).
 DE NXFL.
 GN Coturnix coturnix japonica (Japanese quail).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Coturnix.
 OC NCBI_TaxID=93934;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS OF GLU-246.
 RX MEDLINE=21257892; PubMed=11358864;
 RA Coburn G.A., Wiegand H.L., Kang Y., Ho D.N., Georgiadis M.M., Cullen B.R.;
 RT "Using viral species specificity to define a critical protein/RNA interaction surface.";
 RL Genes Dev. 15:1194-1205(2001).
 CC -!- FUNCTION: Involved in the nuclear export of mRNA from the nucleus to the cytoplasm.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- DOMAIN: The leucine-rich repeats and the NTF2-domain are essential for the export of mRNA from the nucleus (By similarity).
 CC -!- MISCELLANEOUS: Changing a single residue, Glu-246 to Arg, the residue found in human NXFL1, fully rescues both constitutive transport element (CTE) function and CTE binding.
 CC -!- SIMILARITY: BELONGS TO THE NXF FAMILY.
 CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).
 CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -!- SIMILARITY: CONTAINS 1 UBA DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF343749; AAK58910.1; -
 CC PROSITE; PS50177; NTF2_DOMAIN; 1.
 CC Transprot; mRNA transport; Nuclear protein; Repeat;
 CC Leucine-rich repeat; Multigene family.
 CC DOMAIN 116 195 RNA-BINDING (RRM) (BY SIMILARITY).
 CC REPEAT 263 288 LRR 1.
 CC REPEAT 289 312 LRR 2.
 CC REPEAT 313 340 LRR 3.
 CC REPEAT 341 368 LRR 4.
 CC DOMAIN 383 533 NTF2.
 CC UBA-LIKE (BY SIMILARITY).
 CC 563

Query Match 3.6%; Score 114; DB 1; Length 870;
Best Local Similarity 20.6%; Pred. No. 1.3;
Matches 121; Conservative 92; Mismatches 195; Indels 180; Gaps

```

QY   3 SFTATAAAGSGRWLGNNHQPPLSSSOSDLS---YCSSLPMASRVTKLNVSSAL--HTP 57
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   384 STATSSSYQK-----QRPTYSKSKNNWTFGRASDTPLPPHATPK-NVDAFVTADTP 437
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   58 PALHF-----PKQSSNSPAIVVKPKAESNTKMNLFORAAAAADAAGEFLVSHE 108
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   438 PAQTETPEVPSTPQSS-----PPTAKEPUSSNL-----PKTPIISIPPTLOQS 485
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   109 KLHPAP-KTADPSV-----QIAGNFAVDNQPVVRNLVVVGKLPDSIKGVYVRNGA 158
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   486 KTKPLUPVFPAGESLSLTATVDNQPSGQ---VDSRLLHIRAPA---LPPSRKQNFIHRD 539
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   159 NPLHPVTVGHFFDDGMVHAKEHGHSAYACRFTOTNRFOERQLGRPv---FPKAIG 215
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   540 SOLYDSLp-----NHGSGA-----TPTSLSISIPOERPVSVLSSQITG 578
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   216 ELHGHTGARMLFYARAAAGIvDP-AHGTVANAGLVYFNGLAMSEDDLPLYQVOITP 274
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   579 ELRE-----LNPGATGSTSLVGOSLFQHSHLSLTSQPLGNASTAEVL 620
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   275 NGDKTVGRFDGDGLESTMAHKVPDESELFSALSY---DVYSKPYLYKFSPDGCK 331
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   621 NASEFK-----DGLMQLNSLI-----GEI-ALNYLPNSVMNSPL-----PIG-- 655
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   332 SPDVEIQLDQTMHMDFAITENFV-VVPQQVVKFLPEMIRGGSPVVYDNKNKVARFGIL 389
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   656 ---INLRINGAKFEKVILLNQAFIERVAPEE---FKV-----NPSFIDSRTLGAi--- 699
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   390 KYAADSSNIKWIDAPDCFCFH-LWNAREEPETDEVVVGSGMTPTDPSI----- 437
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   700 -KY-----SIKEPTAP-IYLHPWR-FESHQASVULTVKMSLSLEDEISQIVIEDLVVF 750
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   438 FNESDENLKSVLSIRNLKTGEST-----RRPII---SNEDQQV---NLEAGMVNRNML 486
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   751 VNIDGANATALSPOGSFSFKRKRIITWRKEPWLTRNGEGORLIARTITDLAHESAK 810
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
QY   487 GRKTKFAYALAE----PWPKVSGFAKVDLTTCGEVKKHLYGDNRYGGE 530
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||
Db   811 GVITKf---ISETIDNALFHSGSGGITTCOELDE---NNPFGE 851
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :||

RESULT 3
DAAA_PROMI
ID DAAA_PROMI STANDARD; PRT; 466 AA.
AC P28837;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chromosomal replication initiator protein dnaA.
GN DNAA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Proteus.
QC NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LMI509;
RA MEDLINE=91033012; PubMed=2172087;
RX Skovgaard O.;
RT "Nucleotide sequence of a proteus mirabilis DNA fragment homologous
RT to the 60k-rnpA-rpmH-dnaA-dnan-recF-gyrB region of Escherichia
RT coli.";
RL Gene 93:27-34(1990).
CC CC
CC CC
CC CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;
CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS
CC (DNAA BOX): 5'-TTATC(C/A)(C/A)-3'. DNAA BINDS TO ATP AND TO
CC ACIDIC PHOSPHOLIPIDS.
CC -! SIMILARITY: BELONGS TO THE DNAA FAMILY.
```

```

CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: M58352; AAA83958.1; -.
CC  PIR: J00733; IOEBV.
CC  InterPro: IPR001957; Bac_DnaA.
CC  Pfam: PF00308; bac_DnaA; 1.
CC  PRINTS: PR00051; DNaA.
CC  TIGRfam: TIGR00362; DnaA; 1.
CC  PROSITE: PS01008; DNaA; 1.
CC  DNA replication; DNA-binding; ATP-binding.
CC  NP_BIND 171 178 ATP (POTENTIAL).
CC  FT BIND 171 178
CC  SQ SEQUENCE 466 AA; 52974 MW; F8B67C142FE9FA41 CRC64;
CC  -----
CC  Query Match          3.6%; Score 113; DB 1; Length 466;
CC  Best Local Similarity 21.9%; Pred. No. 0.64;
CC  Matches 84; Conservative 48; Mismatches 112; Indels 140; Gaps 21;
CC  QY 105 VSHEKLHLPKT-ADPSVOIAGNAPVNEQPVVRNLPVVGKLPDS-IGKVYVRNGANPLH 162
CC  Db 83 VSARTIESVPKTVTHPAV-----NSTFTNSQPVYR---PSWDNQPSQLPELNYSNVNPKH 135
CC  QY 163 EPVTGHHFFDGDGMVHAKFEHGSASYACRETQTNRFCVEROLG--RPVFPKAIQELGH 220
CC  Db 136 K-----FD-----NFVEGKNQQLAR--AAARQVADNPGGAYNPLF-----LYGG 172
CC  QY 221 TGIAKMLFYARAAAGIVDPAGTGV-----ANAGLVYFNGRLLAMSEDDLPYQVQITPNG 276
CC  Db 173 TGLGTHLLH-----AVGNSIMERKANAKVVMYH-----SERVQDMVKALQNN 216
CC  QY 277 DLKTVGRDFDQGLRSTMIAPKVPDESGELFALSDYVVVSKPLKPYFRPSDGTSPDVE 336
CC  Db 217 ALE-----DFK-----RYYR-SVDALLIDDIQ 237
CC  QY 337 IQLDQTMHDFAITENFVVPDQOWF---KLPEMIRGSPVVDKKNVAREGILDXYA 393
CC  Db 238 PFANKERSQEEFFHFNALLENQIILTSRYPEING-----VEDRLK-SRFG-----286
CC  QY 334 EDSSNIKIDAPDCFCFHLNWAEEPTDEVVIGSCMTTPDPSIFENSDENLKSVLSEI- 452
CC  Db 287 -----WGLTVAIEPPELETRVAI-----LMKKADENQIQLPDEVA 321
CC  QY 453 -----RLNLKTGSTRRPIISN 469
CC  Db 322 FFIARLRSNVRELEGALNRVIAN 345
CC  -----
CC  RESULT 4
CC  PAK1_SCHPO
CC  ID PAK1_SCHPO STANDARD; PRT; 658 AA.
CC  AC P50527;
CC  DT 01-OCT-1996 (Rel. 34, Created)
CC  DT 30-JUN-2000 (Rel. 39, Last sequence update)
CC  DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC  DE Serine/threonine-protein kinase pak1/shk1 (EC 2.7.1.-).
CC  GN PAK1 OR SHK1 OR ORB2 OR SPBC1604.14C.
CC  OS Schizosaccharomyces pombe (Fission yeast).
CC  OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC  OC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC  OC Schizosaccharomycetes.
CC  OX NCBU_TaxID=4896;
CC  RN [1]
CC  SEQUENCE FROM N.A.
CC  RX MEDLINE=96112805; PubMed=8846783;
CC  RA Ottilie S., Miller P.J., Johnson D.I., Creasy C.L., Sells M.A.,
CC  RA Bagrodia S., Forsburg S.L., Chernoff J.,
CC  RT "Fission yeast pak1+ encodes a protein kinase that interacts with

```

```

RT Cdc42p and is involved in the control of cell polarity and mating.";
RL EMBL J. 14:5908-5919(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Marcus S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Squoros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Houry S., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weijens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [4]
RP SEQUENCE OF 119-658 FROM N.A.
RX MEDLINE=95320235; PubMed=7597098;
RA Marcus S., Polverino A., Chang E., Robbins D., Cobb M.H.,
RA Wigler M.;
RT "SKL1, a homolog of the Saccharomyces cerevisiae Ste20 and mammalian
RT p66PAK protein kinases, is a component of a Ras/Cdc42 signaling
RT module in the fission yeast Schizosaccharomyces pombe.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:6180-6184(1995).
CC 1- FUNCTION: FORMS AN ACTIVATED COMPLEX WITH GTP-BOUND RAS-LIKE
CC CDC42. ACTS IN SIGNAL TRANSDUCTION. INVOLVED IN THE CONTROL OF
CC CELL POLARITY AND MATING. MAY INTERACT WITH BYR2.
CC 1- PTM: AUTOPHOSPHORYLATED ON SERINE RESIDUES.
CC 1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC 1- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U22371; AAC49125.1; -.
CC EMBL: AL034433; CAA22347.1; -.
CC EMBL: L41552; AAB52609.1; -.
CC HSSP: Q00534; 1B18.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR000095; PAKbox/Rhodog.
CC InterPro: IPR002290; Ser_thr_kinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam: PF00005; pkinase; 1.
CC Pfam: PF00786; PKD; 1.
CC PRINTS: PR00109; TYRKINASE.
CC PRODOM: PD0000001; Euk_pkinase; 1.

```

```

[1]
SEQUENCE FROM N.A.
RC STRAIN-White leghorn; TISSUE=Limb bud;
RA MEDLINE=93300846; PubMed=114802;
RX Shinomura T., Nishida Y., Ito K., Kinata K.;
RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT expressed during chondrogenesis in chick limb buds. Alternative
RT spliced multiforms of PG-M and their relationships to versican.";
RL J. Biol. Chem. 268:14461-14469(1993).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 2 isoforms: V0 (shown here) and
CC V1; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC developing limb buds.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60226; CAA42787.1; -;
DR EMBL; D13542; BAA02742.1; -;
DR HSPD; P00740; IEDM.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig_MHC.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_Ca; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_Ca; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
ET SIGNAL 1 26
ET CHAIN 27 3562
ET DOMAIN 37 136
ET DOMAIN 166 243
ET DOMAIN 264 345
ET DOMAIN 264 345
ET LINK 2

```

```

FT DOMAIN 3254 3290 EGF-LIKE 1.
FT DOMAIN 3292 3328 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3341 3355 C-TYPE LECTIN.
FT DOMAIN 3460 3518 SUSH1.
FT DISULFID 44 129 BY SIMILARITY.
FT DISULFID 171 242 BY SIMILARITY.
FT DISULFID 195 216 BY SIMILARITY.
FT DISULFID 269 344 BY SIMILARITY.
FT DISULFID 293 314 BY SIMILARITY.
FT DISULFID 3258 3269 BY SIMILARITY.
FT DISULFID 3263 3278 BY SIMILARITY.
FT DISULFID 3280 3289 BY SIMILARITY.
FT DISULFID 3296 3307 BY SIMILARITY.
FT DISULFID 3301 3316 BY SIMILARITY.
FT DISULFID 3318 3327 BY SIMILARITY.
FT DISULFID 3334 3345 BY SIMILARITY.
FT DISULFID 3362 3454 BY SIMILARITY.
FT DISULFID 3430 3446 BY SIMILARITY.
FT DISULFID 3461 3504 BY SIMILARITY.
FT DISULFID 3490 3517 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 948 948 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 485 1411 MISSING (IN ISOFORM V1).
SQ SEQUENCE 3562 AA; 388078 MW; 9BC5656E880C1602D2 CRC64;

Query Match 3.4%; Score 106; DB 1; Length 3562;
Best Local Similarity 22.7%; Pred. No. 41;
Matches 118; Conservative 61; Mismatches 191; Indels 150; Gaps 29;

QY 1 MASETATA-AVSGRWLGGNHGHTOPPLSSGSSDLSYCSLPMASRVTRKLNVSALHT--- 56
DB 1231 MLSSFTAGSILLTLGASPSQTPGSGISSELEEVKIVPFSRAUDKTVTISLDTSSI 1290

QY 57 -----PPALHFPKSSNSPAIVKPKAKESNTKQMNLI-----FQRAAAALDAAEGL 104
DB 1291 SAVDKIQPTASKEPVSSKSPRII--PEBDEVTSSDIIVIDESISPSKASAEDDLTKGM 1348

QY 105 VSHEKLHLPK-----TADPSVQIAGN-FAPVNEQPVRRNLPVVGKLPDS- 148
DB 1349 VEPE-----IDKEYFTSSTATAVARTAPTVMTEATEALQPVSVPTSH-----PDSG 1396

QY 149 --IKGVYVRNGANPLHPEVTGHH-HEFDGDMVHAKFHEGHSASYACRFQTQTRFVQERQ- 204
DB 1397 TDIRLYVIGITGNDTHDPNFEILDLSRHLPLPHANDETHDAESAGTECTSDSVQDSSE 1456

QY 205 --LGRVPFPAIGELHGHGTGIALRLMLFYARAAAGIVDPAHGTCVAN-AGLVYFNGRLLAM 261
DB 1457 YIILODPFF-----NFMDFEIEEE-----DCENTTDTVTTPPALQFNGK----- 1495

QY 262 SEDDLPLQVQITPBGDLKTVGFRDFDQGLSTHIAHPKVDPSGELFALSYDWSKPYLK 321
DB 1457 YIILODPFF-----NFMDFEIEEE-----DCENTTDTVTTPPALQFNGK----- 1495

Db 1496 -----QQVTSAPKSKAEARSD---QIES--VAHSK-NVTFSQLNTEFTIIST--- 1540
QY 322 YRFPSPDGTKSPDV-----ETQLDQPTMMHDFAITENVVVPDQOVVKPEMIRGGSP 375
Db 1541 ----EASGTMQPSKAGVMGAFEVTPQTA--DVAMLE-----P 1572
QY 376 VYDKNKVARFGILDKYAEKSS-----NIKWDAPDFCFHNLWANE--PETDEV 424
Db 1573 VYSGEVEV---TTDKYLEITSVYEQSPKKNKTV-----MMHGTFESSTKTKNL 1620
QY 425 VVIGSCMTPPDSIFNESDENLKSILSEIRLNKLTGSTRR 464
Db 1621 LLITNESSGDGS--TESDLS-RSVFTEI-LTMSSHEDSEK 1656

RESULT 6
CAD5_MOUSE STANDARD; PRT; 783 AA.
ID CAD5_MOUSE
AC P55284; O35542;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Vascular endothelial-cadherin precursor (VE-cadherin) (Cadherin-5).
OS CDH5.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain capillary;
RX MEDLINE=96141083; PubMed=8555485;
RA Breier G., Breviaro F., Caveda L., Berthier R., Schnuerch H.,
RA Gotsch U., Vestweber D., Risau W., Dejano E.;
RT "Molecular cloning and expression of murine vascular endothelial-
RT cadherin in early stage development of cardiovascular system.";
RL Blood 87:630-641(1996).
RN [2]
RP SEQUENCE FROM N.A. AND FUNCTION.
RC STRAIN=BA6/c; TISSUE=Breast carcinoma;
RX MEDLINE=97364256; PubMed=9220534;
RA Matsuyoshi N., Toda K.-I., Horiguchi Y., Tanaka T., Nakagawa S.,
RA Takeichi M., Imamura S.;
RT "In vivo evidence of the critical role of cadherin-5 in murine
RT vascular integrity.";
RL Proc. Assoc. Am. Physicians 109:362-371(1997).
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS.
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X83930; BAA58782.1; -
CC EMBL: D63942; BAA22617.1; -
CC HSSP: P15116; INCU.
CC MGD: MGI:105057; Cdh5.
CC InterPro: IPR002126; Cadherin.
CC Pfam: PF00028; cadherin; 5.
CC Pfam: PF01049; Cadherin_C_term; 1.
CC PRINTS: PR00205; CADHERIN.
CC SMART: SM00112; CA; 5.
```

PROSITE; PS00232; CADHERIN_1; 3.
 Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 Signal.

FT FT SIGNAL 1 24 POTENTIAL.
 FT FT PROPEP 25 45 POTENTIAL.
 FT FT CHAIN 46 783 VASCULAR ENDOTHELIAL-CADHERIN.
 FT FT DOMAIN 46 592 EXTRACELLULAR (POTENTIAL).
 FT FT TRANSMEM 593 619 POTENTIAL.
 FT FT DOMAIN 620 783 CYTOPLASMIC (POTENTIAL).
 FT FT DOMAIN 46 148 CADHERIN 1.
 FT FT DOMAIN 149 255 CADHERIN 2.
 FT FT DOMAIN 256 370 CADHERIN 3.
 FT FT DOMAIN 371 475 CADHERIN 4.
 FT FT DOMAIN 476 592 CADHERIN 5.
 FT FT DOMAIN 737 752 SER-RICH.
 FT FT CARBOHYD 59 59 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT FT CARBOHYD 154 154 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT FT CARBOHYD 440 440 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT FT CONFLICT 67 69 KDO -> GKIK (IN REF. 2).
 SQ SEQUENCE 783 AA; 87847 MW; 240AF2D663BCE7IC CRC64;

Query Match
 Best Local Similarity 17.7%; Score 105; DB 1; Length 783;
 Matches 128; Conservative 101; Mismatches 242; Indels 254; Gaps

QY 60 LHPFKQSSNSPAIVVKKAKESNTKQNLQFQRAAA--ALDAAEGFLYSHEKLHP---- 112
 Db 52 MHIDEENESLPHVYKQDSNVNRQNAKYVIGGFAGKIFGVDDATNGVUAYERLDRKYS 111
 QY 113 -----LPKTADPSVQ-----TAGFAPVNEQPVRRNLFPVVGKLPDSIKGVY 153
 Db 112 EYFLTALIVDKNTKNLEQSPSFTVKVHDINDNWPVESHQVFNASVPMSAIGTSIRVT 171
 QY 154 VRNGANP-----LHEPYTGHFF--DQDGMVHAVKFEHGSYACRPTQTNRVQER 203
 Db 172 AVDADDEFTVGHATVLQIVKGNIEYSIDNSGLI-----FTKIKNLDREK 216
 QY 204 QLGRPVFPK--AIGEILHGTGLARMLFYARAAAGIVDPAHGTGVANAGLVYENGLLA 260
 Db 217 QAQYKIVVETQDALG-LRGESEGTAVMI-----RL-- 245
 QY 261 MSEDDLFPYQVITPNGDLKTVGRFDGQLESTMAHP-----KVDPSEGLFALSYDV 315
 Db 246 --ED-----INDNFVFTQSTYTESVP-EDIRVGKPLGFLRVDPDPQNRRTKYSIM 295
 QY 316 SKPKLYKFRFSPD-----GTKSPDVE-----IQLODPTKMHDP----- 348
 Db 296 QGEYRDYETETDPKRNEGIIKPKSLDYEVIQYTFYIETDPTIRYELSLSTSGKNKA 355
 QY 349 AITENFVVVDQVW-----FKLPEMIRG---GSPVVYDKNKVAR-FGLDKYAESDS- 397
 Db 356 MVTINVLVDVDPVPFQRHFHFKLPEKQKPLGTVVAKDPDKAQRSIGYSIRKTSDRGQ 415
 QY 398 -----NIKWIDAPDCFPLHNAWEEPTDE-----VVVIGSCMT 432
 Db 416 FFRITKQGNIIYNEKELDRETYAWNL--TVEANELSRGNPVGKESIVQYIEVLDDNN 473
 QY 433 PPSDI-----FNSEDNLKSVLSEIRLNKLGESTRRPIISNEDQQ 473
 Db 474 PPEFAQYEPKVCENAAQGLVQISATDKDVVPVNPKEFKALK-NEDSNFTLINNHDT 532
 QY 474 VNL--EAGMVNRNMLGRTKFAVALA---EPKPKVSGFAKVDLTGGEVKKHLYGNRGY 528
 Db 533 ANITVKYQGQNR---EHAKFHYLPVLISDNGVPSLTGTS--FLTGVGCKNCEQGEFTFC 586
 QY 529 GEPLFLPEGGEDEGVILCFV-----HDE 553
 Db 587 EEMAQAQGVSIQALVALFCLILITVTLIIIRRIKQAHAKHSALIEHQLVYTD 646
 QY 554 K-----TWKSEIQIVNAV-----SLEVEATKPLSPVPVGFCHGTFIGADO 593

```

CC DR EMBL: U26644; AAC50259.1;
DR Genew: HGNC:3594; FASN.
DR MIM: 600212;
DR InterPro: IPR001227; Ac_transferase.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; ketoacyl-synt.
DR InterPro: IPR003880; Ppantne_attach.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF001107; adh_zinc; 1.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF00550; pp-binding; 1.
DR Pfam: PF00698; Acyl_transfer; 1.
DR Pfam: PF00975; Thioesterase; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS00075; ACP_DOMAIN; 1.
KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW Pyridoxal phosphate.
FT DOMAIN 1 413
FT DOMAIN 428 815
FT DOMAIN 1630 1857
FT DOMAIN 1858 2113
FT DOMAIN 2118 2174
FT DOMAIN 2202 2504
FT ACT_SITE 161
FT ACT_SITE 580
FT ACT_SITE 876
FT NP_BIND 1666 1683
FT BINDING 1699
FT NP_BIND 1880 1895
FT BINDING 2151
FT ACT_SITE 2302
FT ACT_SITE 2475
FT SEQUENCE 2504 AA; 273100 MW; 8AAF9809B2338DFA CRC64;

Query Match
Best Local Similarity 3.3%; Score 105; DB 1; Length 2504;
Matches 81; Conservative 41; Mismatches 143; Indels 82; Gaps 19;

QY 11 SGRWLGNNHTQPTLLSSQSSDLSYCSLPLWASRVTKLN--VSSALHTPPALHPKQS-- 66
Db 707 SARWL-----STSPDAQHSSSLARTSSAEYNNVNNLSPVLFQFALWVPEHAVV 756
QY 67 -----SNSPAIVVVKAKESNTKQMNLFQRAAAALDAAGFLVSHEKHLPLKPTADPSV 121
Db 757 LEIATPTCPQAVLKVRKPSCT-----IIPRMKKDHRDNEFFLAGTGRHLSCIDANPA 812
QY 122 QIAGNFAPVNEQPVRRNLPPVCKLPDSIKGVYVRGANGPLHEPVTGHHFFDGDGMVHAYK 181
Db 813 L-----FPPV-ESPARGTPLISPL-----IKWDHSLWADAPA-AED 847
QY 182 FHGSAS-----YAC--RFTQNRFRVQERQL--GRPVFEPKAGELGHTGHTGIARMLFYARA 233
Db 848 FPNQSGSPSATIYCTPSSSDPRVINDTIDGRVLFPAI-----GYSIVWKTLL--ARA 900
QY 234 AGI-----VDPAGTGVANAGLVYNGRLLAMSEDDLPYQVQITPNGDGLKTVGR- 283
Db 901 WAGLEQLPVVFDVVQHQATILPKTGTVTSLEVRL-----EAIGAFVSENGNLVSGKV 955
QY 284 FDFDQOLESTMTAHPK-VDPES--GELFALSVDVYKPKYKLFERFSP 327
Db 956 YQWDDP-DPRFDHPSPHPNPSRSPFLFQAQAEVYKELRLRGDYGP 1001

RESULT 8
ID APCE_FREDI
AC P16566;
DT 01-AUG-1990 (Rel. 15, Created)

```

```

DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 01-DEC-1992 (Rel. 24, Last annotation update)
DE Phycobillosome 120 kDa linker polypeptide, core (L-CM 92) (Core-
DE membrane linker protein).
GN APCE.
OS Fremyella diplosiphon (Calothrix PCC 7601).
OC Bacteria; Cyanobacteria; Nostocales; Rivulariaceae; Fremyella.
CX NCBI_TaxID=1197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192765; PubMed=2107546;
RA Houmad J., Capuano V., Colombano M.V., Coursin T., de Marsac N.;
RT "Molecular characterization of the terminal energy acceptor of
RT cyanobacterial phycobillosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2152-2156(1990).
CC -1- FUNCTION: THIS PROTEIN IS POSTULATED TO ACT BOTH AS TERMINAL
CC ENERGY ACCEPTOR (BY ITS PHYCOBILIN-LIKE DOMAINS) AND AS A LINKER
CC POLYPEPTIDE (BY ITS REPEATS AND ARMS) THAT STABILIZES THE
CC PHYCOBILISOME CORE ARCHITECTURE.
CC -1- SUBUNIT: PHYCOBILISOMES OF THIS ORGANISM ARE COMPOSED OF A TWO
CC CYLINDER CORE, FROM WHICH SIX RODS RADIATE. THE CORE IS MAINLY
CC COMPOSED OF ALLOPHYCOCYANIN ALPHA AND BETA CHAINS, AND OF THREE
CC MINOR COMPONENTS: THE ALLOPHYCOCYANIN ALPHA-B CHAIN, A 18.3 kDa
CC POLYPEPTIDE, AND THE ANCHOR POLYPEPTIDE LCM.
CC -1- SUBCELLULAR LOCATION: ANCHORS THE PHYCOBILISOME PERPENDICULARLY
CC TO THE STROMAL SURFACE OF THE THYLAKOID MEMBRANE.
CC -1- SIMILARITY: THE REPEATED DOMAINS ARE SIMILAR TO THE N-TERMINAL
CC REGIONS OF PHYCOCYANIN ROD LINKER POLYPEPTIDES.
CC -1- SIMILARITY: THE PHYCOBILIN-LIKE DOMAINS ARE SIMILAR TO PHYCOBILINS
CC FROM VARIOUS SPECIES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M20806; AAC24873.1;
DR PIR: A35088; A35088.
DR HSP: P00318; 1B33.
DR InterPro: IPR001297; PBS_linker_poly.
DR InterPro: IPR001659; Phycobillosome.
DR Pfam: PF00427; PBS_linker_poly; 4.
DR Pfam: PF00502; Phycobillosome; 1.
DR ProDom: P000340; Phycobillosome; 2.
KW Phycobillosome; Electron transport; Photosynthesis; Repeat.
FT INIT_MET 0 0
FT DOMAIN 17 75 PHYCOBILIN-LIKE 1.
FT DOMAIN 76 143 PHYCOBILIN-LIKE LOOP.
FT DOMAIN 144 236 PHYCOBILIN-LIKE 2.
FT DOMAIN 237 284 ARM 1 (SPACING SEQUENCE).
FT REPEAT 285 409 I.
FT DOMAIN 410 546 ARM 2 (SPACING SEQUENCE).
FT REPEAT 547 669 II.
FT DOMAIN 670 743 ARM 3 (SPACING SEQUENCE).
FT REPEAT 744 869 III.
FT DOMAIN 870 953 ARM 4 (SPACING SEQUENCE).
FT REPEAT 954 1079 IV.
FT SEQUENCE 1079 AA; 120325 MW; 87FE38F232BF82 CRC64;

```

```

Query Match
Best Local Similarity 3.3%; Score 104.5; DB 1; Length 1079;
Matches 71; Conservative 47; Mismatches 80; Indels 135; Gaps 21;

QY 62 FPKQSSN---SPAIVVVKPAKESNTKQMNLFQRAAAALDAAGFLVSHEKHLHP-----L 113
Db 467 FPKETRNPTSPA---PFSKTRRLIN-----OGPGINSQVSNPGARGE 509
QY 114 PKTADPSV-----QIAG-----NFAPVNEQPVRRNLPPVVKLPDSIKGVYVRNG 157
Db 510 PGSGLPKVFRDLQPLGTIGKKAAGKASIKFSESSQAV-----IKAYLQ-- 554

```

QY	158	ANPLHEPVTHGHHFDGDM-VHNVFHEGHSAS-----YACRET	194
DB	555	-----VFGROVYEGQRLKVOEIKLENGQLSVREFIRALAKSDVERKYTWTSLSYVCKAI	607
QY	195	QTNRFVQERQLGRVPFKPAIGELHGHGTGLARLMLFYARAAAGIVDPAHGTGVANAGLVYF	254
DB	608	E---YHRRLLGRPTVGRQ--EINKYFDAAKGGFYA-----VVD-----AIIINS--VEY	650
QY	255	NGRLAMSEDDLPYQVOITPNG-----DLKTVG--RDFDQGLESTMIAPKVPDPESGELF	308
DB	651	SE---AFGEDTVPYERYLTSPGVALQRLR-VGSIREDVGS-----KVOKQETPLF	696
QY	309	ALSYDVVSRPKLYKFRFSPDGTSPDVEIQLDQ	341
DB	697	VTLGTVT-----DTRTPDQIQRINQ	717
RESULT 9			
ID	SYFB_AQUAE	STANDARD;	PRT; 775 AA.
AC	067620;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--		
DE	tRNA ligase beta chain) (PheRS).		
GN	PHE1 OR AQ_1730.		
OS	Aquifex aeolicus.		
OC	Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;		
OC	Aquifex.		
OX	NCBI_TaxID=63363;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-VF5;		
RA	Medline=98196666; PubMed=9537320;		
RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,		
RA	Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,		
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,		
RT	"the complete genome of the hyperthermophilic bacterium Aquifex		
RT	aeolicus."		
RT	Nature 392:353-358(1998).		
CC	-!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +		
CC	diphosphate + L-phenylalanyl-tRNA(Phe).		
CC	-!- SUBUNIT: Tetramer of two alpha and two beta chains (BY		
CC	SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA		
CC	CHAIN FAMILY. SUBFAMILY 1.		

This SWISS-PROT entry is copyright. It is produced through a collaboration			
between the Swiss Institute of Bioinformatics and the EMBL outstation -			
the European Bioinformatics Institute. There are no restrictions on its			
use by non-profit institutions as long as its content is in no way			
modified and this statement is not removed. Usage by and for commercial			
entities requires a license agreement (See http://www.isb-sib.ch/announce/			
or send an email to license@isb-sib.ch).			

DR	EMBL; AF000753; AAC07582.1;		
DR	HSSP; P27002; 1PYS.		
DR	InterPro; IPR005146; B3_4.		
DR	InterPro; IPR005147; B5.		
DR	InterPro; IPR005121; Fdx-Antich.		
DR	InterPro; IPR004532; PheT_Bact.		
DR	InterPro; IPR002547; tRNA_bind.		
DR	Pfam; PF01588; tRNA_bind; 1.		
DR	Pfam; PF03147; FDX-ACB; 1.		
DR	Pfam; PF03483; B3_4; 1.		
DR	Pfam; PF03484; B5; 1.		
DR	TIGRfam; TIGR00472; pheT_bact; 2.		
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; AIP-binding.		
KW	Complete proteome.		
CC	SEQUENCE 775 AA; 88484 MW; 61BB15FDE5B274C8 CRC64;		

Query Match			
3.3%; Score 103; DB 1; Length 775;			
Best Local Similarity 20.7%; Pred. No. 7.4;			
Matches 114; Conservative 67; Mismatches 159; Indels 212; Gaps 31;			
QY	123	IAGNAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPLHEPVTHGHHFD	172
DB	46	VFGKVVEVKEHTKKKLAVV-----KVQVEHIFIDVVTVKSVRE	86
QY	173	GGGMVHAKFEHGSASACRETQTNRFVQERQLGRVPFKPAIGELHGHGTGLARLMLFYAR	232
DB	87	GGCVI--VALPNKAVGNMC-----VTEREF-----DGVSCKGLLSA	121
QY	233	AAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLP	277
DB	122	QELGLEEKSEG-----VLKIHEDFKPGTDANEILGFGKEKIIEDITPNRGD	167
QY	278	LKTVGREFDQGLESTMIAPKVPDP-----ESGELFALSYDVVSRPKLYKFRFSPDGTK	331
DB	168	MLSVRGVARD---LSAIFRLPKKKPEPTYEETGETFEIEDCKKY-----RGVV	216
QY	332	SPDVEIQLDQ-----TMMHDFAITENFVVVDPQVVFVKLPPEMIRGSPV-VY	378
DB	217	IEGVEIK-ESPLYKKRLWQGGIKSINNVDITNV-----MLRDGQPLHAF	262
QY	379	DNKNVAREGILDYAESSNKKWIDAPDCFCFHLNNAWEPEPDEVVVGISCMTPPDSIF	438
DB	263	DLSKV-EGGIIVRSAKKGEKIITLDG-----EERELEDILVIAADREKPLAVA	309
QY	439	N-----ES--DENLKSVLSEI-----RLNLKTGSTRPRPIISNEDQOVNLEA	478
DB	310	GVIGLGSEGKENTKOILLESAYENPFRVRKASKKLGIOLESSYR-----FERNVDIE-	362
QY	479	GMVNRNMLGRKTKFA-YIALAEPKPVSGPAKVDLTGTGEVKKHLYGDNRYGGEPLFLP--	535
DB	363	-----RVDRADQYAVYLLKHAGKV-----KVKQVYRE-KYKPKKVELPQG	404
QY	536	-----GEGGEDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVPYGFHGTFF-	588
DB	405	KYIRYAGESYKNEE-----VKELIDALEIPNEI-MRCGVEVLVFS-----HRSFD	448
QY	589	IGAD-DIAKQVV	599
DB	449	IQRVDLIEEM	460
RESULT 10			
ID	CYAA_USTMA	STANDARD;	PRT; 2493 AA.
AC	P49606;		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl		
DE	cyclase).		
GN	UAC1 OR REM1		
OS	Ustilago maydis (Smut fungus).		
OC	Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;		
OC	Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.		
OX	NCBI_TaxID=5270;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-518;		
RA	MEDLINE=95087882; PubMed=7995519;		
RA	Gold S., Duncan G., Barrett K., Kronstad J.W.;		
RT	"cAMP regulates morphogenesis in the fungal pathogen Ustilago		
RT	maydis."		
RL	Genes Dev. 8:2805-2816(1994).		
CC	-!- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR		
CC	METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,		
CC	CAMP.		
CC	-!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.		
CC	-!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.		

```
CC CC -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC CC -!- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC CC at the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; L33918; AAA57469.1; -.
DR DR InterPro; IPR001054; G_Cyclase.
DR DR InterPro; IPR001611; LRR.
DR DR InterPro; IPR003592; LRR_out.
DR DR InterPro; IPR003591; LRR_typ.
DR DR InterPro; IPR001932; PP2C-like.
DR DR InterPro; IPR000159; RA_domain.
DR DR Pfam; PF00211; guanylate_cyc; 1.
DR DR Pfam; PF00481; pp2c; 1.
DR DR Pfam; PF00560; LRR; 17.
DR DR SMART; SM00044; CYCc; 1.
DR DR SMART; SM00370; LRR; 6.
DR DR SMART; SM00369; LRR_TYP; 2.
DR DR SMART; SM00332; PP2C; 1.
DR DR SMART; SM00314; RA; 1.
DR DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.
FT DOMAIN 759 763 POLY-ASP.
FT DOMAIN 888 897 POLY-ALA.
FT REPEAT 1084 1106 LRR 1.
FT REPEAT 1108 1131 LRR 2.
FT REPEAT 1132 1155 LRR 3.
FT REPEAT 1157 1178 LRR 4.
FT REPEAT 1179 1201 LRR 5.
FT REPEAT 1202 1225 LRR 6.
FT REPEAT 1227 1249 LRR 7.
FT REPEAT 1249 1271 LRR 8.
FT REPEAT 1273 1295 LRR 9.
FT REPEAT 1314 1337 LRR 10.
FT REPEAT 1338 1360 LRR 11.
FT REPEAT 1361 1384 LRR 12.
FT REPEAT 1386 1407 LRR 13.
FT REPEAT 1408 1430 LRR 14.
FT REPEAT 1432 1455 LRR 15.
FT REPEAT 1509 1529 LRR 16.
FT REPEAT 1533 1556 LRR 17.
FT REPEAT 1557 1580 LRR 18.
FT REPEAT 1581 1604 LRR 19.
FT REPEAT 1606 1628 LRR 20.
FT REPEAT 1633 1656 LRR 21.
FT REPEAT 1722 2001 PP2C-LIKE.
FT DOMAIN 2002 2493 CATALYTIC.
SQ SEQUENCE 2493 AA; 271979 MW; 106A872C31C5B8F8 CRC64;

Query Match          3.2%; Score 102; DB 1; Length 2493;
Best Local Similarity 21.7%; Pred. No. 48;
Matches      85; Conservative 44; Mismatches 136; Indels 126; Gaps 17;

QY    4   FTAAVSGRWLGNNIT-----QPPLSSQSDDLSCSLPMASRVTRKLNVSALHTP 57
Db     |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
801   FFKTASLAGSRRTDSDVDPLTPALPLPGSKSVDEAAANKVDVLQQ-TNNAQSALVQQ- 858
QY     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
58   PALHFPKCSSNSPAIVVKPKAKESNTIKQNMLQFAAAAALDAAGELVSH-----SKLHPL 113
Db     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
859   ---QSOSQHHPQSPNVRPSTRGGGAHM--FASAGASAAAAAAGKLGLHRPSKRMNR 913
QY     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
114   PKTAD-----PSVQIAGNFAPVNEQPVRRNLPPVGKLPDSIKGVYVRGANGPLEHPVTG 167
Db     |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
914   PNTAGSVGATRPSTTTTLGSTLSAEDD-----TSINGSTRDG-HPLKRSATA 959
QY     |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
168   -----HHFFDGDGMVAHVKEHGASAYACRFTQTINRFVQERQLGRVPFKATGE 216
```


Db 168 AQQSPENPWFQGTADAVRQYLWLFEEHNVLVLAGDHLXYRMDYR-----FIQAHR 221
 Qy 216 ELHGHTGIARLMLFVARAAGIVDPAHGTGVANAGLVFN--GRLLAMSEDDL PYQVQIT 273
 Db 222 ESDADITVAALPMDEARATA-----FCGLKIDESGRILFSEN----- 259
 Qy 274 PNGD-LK-----IVGRDFDQGLBSTMIA-----HPKVD----- 301
 Db 260 PKGOLKANKVDTTILGLDDDDRAKEMPHYIASMGIVYWSKHWMLDLLRDKFPGANDFGSEV 319
 Qy 302 -PESGEL-----FALSVD-----VVSKP---YLKYFRFSPDGTKSPDV 335
 Db 320 IPGATELCMRQVAYLDGYWEDIGTIEAFYNANLIGITKKPVPDFSFYDRSSPIYTQP--- 376
 Qy 336 EIQLDQPTMHDFAITENFV-----VVPDQV---VFKLPEMI----- 370
 Db 377 --RYLPPSKMLDADITDSVIGEGCVIKNCKTHHSVGLRSCISEGAIITLLMGADYYE 434
 Qy 371 -----RGGSPVVYDKNKVAFRGILDKYAEDUSSNIKWDADPCFCHLWNAWEPPE 420
 Db 435 TDADRLAAGKGVFIGIGKNSHRRALIDKNARIGDDVKIINSD-----NVQEAARETE 489
 Qy 421 ----TDEVVVIGSCMTPPDSI 437
 Db 490 GYFIKSGIVTVIKDALIPSGTV 511

 RESULT 14
 ACVS_EMENTI STANDARD; PRT; 3770 AA.
 ID ACVS_EMENTI
 AC P27742;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Delta-(L-alpha-aminoacyl)-L-cysteinyI-D-valine synthetase
 DE (EC 6.-.-.) (ACV synthetase) (ACVS).
 DE ACVA.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 NCBI_taxid:5072;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=G191;
 RX MEDLINE=91286299; PubMed=2061333;
 RA Maccabe A.P., van Liempt H., Pallissa H., Unkles S.E., Riach M.B.R.,
 RA Pfeifer E., von Doehren H., Kinghorn J.R.;
 RA "Delta-(L-alpha-aminoacyl)-L-cysteinyI-D-valine synthetase from
 RT Aspergillus nidulans. Molecular characterization of the acva gene
 RT encoding the first enzyme of the penicillin biosynthetic pathway.";
 RL J. Biol. Chem. 266:12646-12654(1991).
 CC -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
 CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS
 CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
 CC INTERMEDIATES.
 CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETINES
 CC (POTENTIAL).
 CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
 CC CEPHALOSPORIN.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X54853; CAA38631.1; -
 DR

DR PIR: A40889; A40889.
DR HSP: P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; Condensatin.
DR InterPro: IPR003880; Pantine_attach.
DR InterPro: IPR000379; Ser_estrs_site.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 3.
DR Pfam: PF00658; Condensation; 3.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE: PS00455; AMP-BINDING; 3.
DR PROSITE: PS00075; ACP_DOMAIN; 3.
KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW Repeat; Phosphopantetheine.
FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).
FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).
FT DOMAIN 950 919 ACYL CARRIER (ACP) 1.
FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
FT DOMAIN 3020 3087 ACYL CARRIER (ACP) 3.
FT BINDING 882 882 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 1965 1965 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3050 3050 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACT_SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 3770 AA; 422448 MW; CB66B6D232A58CB0 CRC64;

Query Match 3.2%; Score 100; DB 1; Length 3770;
Best Local Similarity 20.9%; Pred. No. 1.2e+02;
Matches 97; Conservative 67; Mismatches 162; Indels 138; Gaps 27;

Qy 12 GRWLGNNHTQPLSSSSDLSYCSLLPMASRVTRK-----INVSAHLHPAL 60
Db 1544 GNWSVSN-----PAPSSTSTDLAY-----IYTSQTGKPKGVVHGVVNNLOISLSKIFGL 1596
Qy 61 -----HFPKQSS-----NSPAIVVKKPKAKESNTKQMLNFORAAAAALDA 99
Db 1597 RDTDEVILSFNVDFHFVQMTDAILNGQTLVNLDMRSKRER--LYQIETNRVY 1654
Qy 100 AGEFLVSHKHLPLKTDAPSVQIAGNAPVNEQPVRR-----NLFVVGKLPDSIKG 151
Db 1655 LSG-----TPSVISMYEFSREKDH-LRRVDCVGEAFSQPVFDQINDTQOG 1698
Qy 152 VYVRNAGNPLHEPVYTGHHFFDGDGMVHAKFEHSASVACRTQNRVQERQLGR-PVF 210
Db 1699 LII-NGYGTPEISITHK-----RLYPPERRTDKSIQGIQNSTSVYLNADMKRVPI- 1750
Qy 211 PKAIGELH-GHTGIARLMFYAR-----AAAGIVDPAGHTGVANAGLVYFNGRLLAISED 265
Db 1751 -GAVGELYLGEGVAR--GYHNRPEVTAERLRNPFQIDSEQRNGR---NSRLYETGD-- 1802
Qy 266 LPYQVOITP--NGDLKTVGRDFD-----GQLESTMIAPH-----KVDPEP 304
Db 1803 ---LVRWTPGSGNETEYLGRLNDFQVIRGLRIELGEIEAVMSHSDPIKQSVVIAKSGREG 1859
Qy 305 GELFALSVDVYVSKPKLYKFRSPDGTSPDVEIQLDQPTMHDFALTENFV-----VYP 358
Db 1860 DQKFLVGEVASSP-----LSPGA-----IRRFMSRLPGYMPSPFIPISLPTVP 1906
Qy 359 DOQVYFK-LPEM-IRGGSPPVYDKNKKVARF-----GILDKYAE 394
Db 1907 SKLDTKALPTAEKGMNVLAAPRNEIESILGIGISALDLSAQ 1950

RESULT 15
ID PHYC_BACSU STANDARD; PRT; 383 AA.
AC Q31097;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

3-phytase precursor (EC 3.1.1.3.8) (Phytate 3-phosphatase) (MYO-
inositol-hexaphosphate 3-phosphohydrolase).
PHYC OR PHYB13.
Bacillus subtilis.
Bacteria, Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=1423;
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
STRAIN=VTT-E-68013;
MEDLINE=98268943; PubMed=9603817;
Kerovuo J., Lauraeus M., Nurminen P., Kalkkinen N., Apajalahti J.;
"Isolation, characterization, molecular gene cloning, and sequencing
of a novel phytase from Bacillus subtilis";
Appl. Environ. Microbiol. 64:2079-2085(1998).
FUNCTION: CATALYZES THE HYDROLYSIS OF INORGANIC ORTHOPHOSPHATE
FROM PHYTATE. ONLY PHYTATE, ADP, AND ATP WERE HYDROLYZED (100, 75,
AND 50% OF THE RELATIVE ACTIVITY, RESPECTIVELY). HAS MAXIMAL
ACTIVITY AT PH 7 AND 55 DEGREES CELSIUS.
CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = D-myo-
inositol 1,2,4,5,6-pentakisphosphate + phosphate.
COFACTOR: ISOLATED ENZYME REQUIRED CALCIUM FOR ITS ACTIVITY AND/OR
STABILITY.
SUBCELLULAR LOCATION: Secreted.
INDUCTION: BY PHYTATE.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

Query Match 3.2%; Score 99.5; DB 1; Length 383;
Best Local Similarity 19.3%; Pred. No. 4.8;
Matches 88; Conservative 71; Mismatches 151; Indels 147; Gaps 22;

Qy 18 NHTQPLSSSSDLSYCSLLPMASRVTRKLN-----VSSALHTPPALHFPKSSNSPA 71
Db 2 NHSKTLTLLTAAGLMLTCCA--VSSQAKHKLSDPVHFTVNAAEETP-VDTAGDAADDDPA 58
Qy 72 IVVRPKAKESNTKQMLNFORAAAAALDAEGLVSHKHLPLKTDAPSVQIAGNAPVN 131
Db 59 IWLDEKTPQ-NSK-----LIITNKKSLGVVYSLDGKMLHSYNTGKLN 99
Qy 132 EQPVERNPPVVGKLPDSIKGVVVRNAGNPLHEPVYTGHHFFDGDGMVHAKFEHSASVAC 191
Db 100 NVDIRVDFPLNGKKVYDIAAASNRSEKNTIE-----IYALDGKNGTL---- 141
Qy 192 RETQNRVQERQLGRPVFP--KAIGELHGTGIARLMFYARAAAGIVDPAGHTGVANA 249
Db 142 -----QSMTPDPHPITATATINEVYGT-----LYHSQK----- 168
Qy 250 GLIVFNGRLAM---SEDDL-PYOVQIITPNDL--KTGREFDQGOLESTMIAPHKVDPE 303
Db 169 -----TGKIYAMVTKGEFEQYELKADKNNGYISKKVYAFKMNQSQTEGM-----AADE 218
Qy 304 SGELFALSVDVYVSKPKLYKFRSPDGTSPDVEIQLDQPTMHDF-----AITENFV 355
Db 219 YGRLYIAFEDEA-----IWKFSAPDGGSGNGTVIDRADGRHLTRQIEGTIYYAAGKGYL 274
Qy 356 VYPDOQVYFKLPEMIRGGSPPVYD---KNK-VAREGILDKVAEDSSNIKIDWADPCFCFH 411

Db 275 MASSQ-----GNSSYAIYDRQGNKYVADFRITDGPETDGT----- 311
Qy 412 LWNWEEPETDEVVIGSCMTP--PDSIFNESD-ENL 445
Db 312 -----DIDGIDLGLGFGFPGIFVQDGENI 340

Search completed: May 19, 2003, 15:48:01
Job time : 45 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:14 ; Search time 43 seconds
(without alignments)
2870.284 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASFTATAVSGRWLGGNHT.....VPGFHGTFIGADDLAKQV 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3150	100.0	599	10 Q9LRR7	Q9LRR7 arabidopsis
2	3143	99.8	599	10 Q93ZU5	Q93ZU5 arabidopsis
3	2289	72.7	604	10 Q9M329	Q9M329 solanum tub
4	2280.5	72.4	605	10 Q24023	Q24023 lycopersico
5	2170.5	68.9	657	10 Q9M9F5	Q9M9F5 arabidopsis
6	2168.5	68.8	612	10 Q9FS24	Q9FS24 vigna ungui
7	2165	68.7	615	10 Q9M6E8	Q9M6E8 phaseolus v
8	2066.5	65.6	625	10 Q9AXZ4	Q9AXZ4 persea amer
9	2044.5	64.9	589	10 Q9C621	Q9C621 arabidopsis
10	1991	63.2	583	10 Q49505	Q49505 arabidopsis
11	1977	62.8	569	10 Q9AXZ3	Q9AXZ3 persea amer
12	1930	61.3	604	10 Q24592	Q24592 zea mays (m
13	1663.5	52.8	577	10 Q9LRR7	Q9LRR7 arabidopsis
14	968	30.7	595	10 Q49675	Q49675 arabidopsis
15	959.5	30.5	543	10 Q94IR2	Q94IR2 phaseolus v
16	938	29.8	538	10 Q9LY63	Q9LY63 arabidopsis

17	937	29.7	538	10	065572	065572 arabidopsis
18	837.5	26.6	200	10	Q94EM8	Q94EM8 lactuca sat
19	798.5	25.3	524	10	Q9AXZ5	Q9AXZ5 persea amer
20	745.5	23.7	475	16	Q8YXV3	Q8YXV3 anabaena sp
21	745	23.7	446	10	Q49895	Q49895 malus domes
22	528.5	16.8	483	16	Q9AA32	Q9AA32 caulobacter
23	527.5	16.7	456	2	Q9AM11	Q9AM11 streptomyce
24	526.5	16.7	456	2	Q93FA4	Q93FA4 streptomyce
25	466	14.8	501	16	Q06785	Q06785 mycobacteri
26	464	14.7	503	16	Q98K46	Q98K46 streptomyce
27	454.5	14.4	502	16	Q05905	Q05905 mycobacteri
28	394	12.5	490	16	P74334	P74334 synchocyst
29	382.5	12.1	209	10	Q8VXN9	Q8VXN9 crocus sati
30	355.5	11.3	497	16	Q8YPB4	Q8YPB4 anabaena sp
31	352	11.2	485	2	Q53353	Q53353 pseudomonas
32	327	10.4	218	10	Q8VXP1	Q8VXP1 crocus sati
33	322.5	10.2	570	10	Q8VY26	Q8VY26 arabidopsis
34	319	10.1	520	16	Q8XTA2	Q8XTA2 ralstonia s
35	317	10.1	480	16	P74370	P74370 synchocyst
36	311.5	9.9	491	2	Q8RQW2	Q8RQW2 pseudomonas
37	303.5	9.6	490	2	Q52008	Q52008 pseudomonas
38	302.5	9.6	552	10	Q93VD5	Q93VD5 oryza sativ
39	270	8.6	342	2	Q68868	Q68868 synchococc
40	260	8.3	526	13	Q91993	Q91993 gallus gall
41	256	8.1	516	13	Q90WH4	Q90WH4 brachydanio
42	253	8.0	616	10	Q9M079	Q9M079 arabidopsis
43	244	7.7	618	10	Q9SHD9	Q9SHD9 arabidopsis
44	242	7.7	185	10	Q8VXP0	Q8VXP0 crocus sati
45	242	7.7	472	16	Q8YMN8	Q8YMN8 anabaena sp

ALIGNMENTS

RESULT 1

Q9LRR7 Q9LRR7 PRELIMINARY; PRT; 599 AA.
AC Q9LRR7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase (Neoxanthin cleavage enzyme).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_taxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty P1 and TAC
clones.";
RL DNA Res. 7:131-135(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-COL;
RA Iuchi S., Kobayashi M., Shinozaki K.;
RT "Characterization of neoxanthin cleavage enzyme from Arabidopsis
thaliana.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028617; BAB01336.1;
DR EMBL; AB026549; BAB70609.1;
DR InterPro; IPR004294; RPB65.
DR Pfam; PF03055; RPB65; 1.
KW Dioxygenase.

SQ SEQUENCE 599 AA; 65856 MW; 7D513F39945E0CF3 CRC64;

Query Match 100.0%; Score 3150; DB 10; Length 599;
 Best Local Similarity 100.0%; Pred. No. 1.8e-237;
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASFTATAAAYSGRWLGNGHHTQPPPLSSQSSDLSYCSLPMASRVTRKLNYSALHTPPAL 60
 DB 1 MASFTATAAAYSGRWLGNGHHTQPPPLSSQSSDLSYCSLPMASRVTRKLNYSALHTPPAL 60

QY 61 HFPKQSSNSPAIVVKKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120
 DB 61 HFPKQSSNSPAIVVKKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120

QY 121 VQIAGNFAPVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 DB 121 VQIAGNFAPVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180

QY 181 KFEHGSASYACRTQTNRFRVQERQOLGRPVPKAIKELHGHGTIARLMLFYARAAAGIVDP 240
 DB 181 KFEHGSASYACRTQTNRFRVQERQOLGRPVPKAIKELHGHGTIARLMLFYARAAAGIVDP 240

QY 241 AHGTGVANAGLVYFNGRLKLLAMSEDDLPYQVQITPNDGLKTVGRFDFDQGLLESTMIAPKV 300
 DB 241 AHGTGVANAGLVYFNGRLKLLAMSEDDLPYQVQITPNDGLKTVGRFDFDQGLLESTMIAPKV 300

QY 301 DPESGELFALSYDVVSKPYLYKFRFSPDGTSPDVEIQDQPTMMHDFATFENFVVVPDQ 360
 DB 301 DPESGELFALSYDVVSKPYLYKFRFSPDGTSPDVEIQDQPTMMHDFATFENFVVVPDQ 360

QY 361 QVVKLPKEMIRGSGSPVYDKNKVARFGLDKYAEEDSNKWKIDAPDCFCFHLNANEPEE 420
 DB 361 QVVKLPKEMIRGSGSPVYDKNKVARFGLDKYAEEDSNKWKIDAPDCFCFHLNANEPEE 420

QY 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPPIISNEDQOVNLEAGM 480
 DB 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPPIISNEDQOVNLEAGM 480

QY 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTVKKHLYGDNRYGGEPLFLPGE 540
 DB 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTVKKHLYGDNRYGGEPLFLPGE 540

QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHTGTFGADDLAKQVV 599
 DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHTGTFGADDLAKQVV 599

RESULT 2

ID Q93205 PRELIMINARY; PRT; 599 AA.

AC Q93205; (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DE Putative 9-cis-epoxycarotenoid dioxygenase.

GN MOA2.4/AT3G14440.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
 RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
 RA Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;

RT "Full length cDNA of gene MOA2.4/AT3G14440 (GI:11994214).";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY056255; AAL07104.1; -
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.
 KW Dioxygenase.
 SQ SEQUENCE 599 AA; 65813 MW; B9007A2DC1C15506 CRC64;

Query Match 99.8%; Score 3143; DB 10; Length 599;
 Best Local Similarity 99.8%; Pred. No. 6.5e-237;
 Matches 599; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASFTATAAAYSGRWLGNGHHTQPPPLSSQSSDLSYCSLPMASRVTRKLNYSALHTPPAL 60
 DB 1 MASFTATAAAYSGRWLGNGHHTQPPPLSSQSSDLSYCSLPMASRVTRKLNYSALHTPPAL 60

QY 61 HFPKQSSNSPAIVVKKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120
 DB 61 HFPKQSSNSPAIVVKKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120

QY 121 VQIAGNFAPVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 DB 121 VQIAGNFAPVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180

QY 181 KFEHGSASYACRTQTNRFRVQERQOLGRPVPKAIKELHGHGTIARLMLFYARAAAGIVDP 240
 DB 181 KFEHGSASYACRTQTNRFRVQERQOLGRPVPKAIKELHGHGTIARLMLFYARAAAGIVDP 240

QY 241 AHGTGVANAGLVYFNGRLKLLAMSEDDLPYQVQITPNDGLKTVGRFDFDQGLLESTMIAPKV 300
 DB 241 AHGTGVANAGLVYFNGRLKLLAMSEDDLPYQVQITPNDGLKTVGRFDFDQGLLESTMIAPKV 300

QY 301 DPESGELFALSYDVVSKPYLYKFRFSPDGTSPDVEIQDQPTMMHDFATFENFVVVPDQ 360
 DB 301 DPESGELFALSYDVVSKPYLYKFRFSPDGTSPDVEIQDQPTMMHDFATFENFVVVPDQ 360

QY 361 QVVKLPKEMIRGSGSPVYDKNKVARFGLDKYAEEDSNKWKIDAPDCFCFHLNANEPEE 420
 DB 361 QVVKLPKEMIRGSGSPVYDKNKVARFGLDKYAEEDSNKWKIDAPDCFCFHLNANEPEE 420

QY 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPPIISNEDQOVNLEAGM 480
 DB 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPPIISNEDQOVNLEAGM 480

QY 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTVKKHLYGDNRYGGEPLFLPGE 540
 DB 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTVKKHLYGDNRYGGEPLFLPGE 540

QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHTGTFGADDLAKQVV 599
 DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHTGTFGADDLAKQVV 599

RESULT 3

ID Q9M329 PRELIMINARY; PRT; 604 AA.

AC Q9M329; (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE Putative 9-cis-epoxycarotenoid dioxygenase.

GN NCED1.

OS Solanum tuberosum (Potato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;

RN [1]

RP SEQUENCE FROM N.A.

RA Burbidge A., Taylor I.B., Thompson A.;

RT "Putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ276244; CAB76920.1; -
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.

Lycopodium esculentum Mill.) library. ;
Exp. Bot. 47:2111-2112(1997).
2]

```

Query Match      68.94;  Score 2170.5;  DB 10;  Length 657;
Best Local Similarity  67.7%;  Pred No. 7,2e-161;
Matches 419;  Conservative 80;  Mismatches 87;  Indels 33;  Gaps 9;

QY      1  MASFTATAAASGRWL-----GGNHTOPPISSSSODLSYCSSLP-----MASRYT 45
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      51  MASTTLLPSTQTQLDRFTSTSSSSRPKIQS-----LSPSTIRNKKLVPCYVSSYN 105

QY      46  RKLNVSSALHW-----PPLHFPKSSNSPALVYKPKAKESNTKMQLMFORAAAAADAAE 101
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      106  KSSVSSLSLQGTFTKPPSW---KKLCNDVTNLI-PKTIINQPK-LNPVQETAAMVLDAVE 160

QY      102  GLVLSHE-KLHPPLKPTADPSVQIAGNFAPVNEQPVRRNLPVGKLPDSTKGVIVRNGANP 160
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      161  NAMISHERRRRPHPKTADPAVQIAGNFPVPPEKTPVWNLHPVTGTVEPICGVTVRNGANP 220

QY      161  LHEPYTGHFFDGDGMHVAVFEGCSASACRFQTQTRFVQEHQOLGRVFPKPAIGELHGH 220
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      221  LHKPVSGHLLFDGDGMHVAVFDFDGSVYACRFTEINRLVOERCCGRPVFPKPAIGELHGH 280

QY      221  TGTARLMLFYARAAAGIVDPAHGTVGANAGVYFNGRLLAMSEDDLPYQVQITPNSDLKT 280
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      281  LGIAKLMLFNTGRGLGVDPDTGGILGVANAGLVYFNHLLAMSEDDLPYHKVQTQSDLET 340

QY      281  VREFPDGGLSTMLAHKVPDPESELEFALSYDVVSKPYLKYFRFSPDCTKSPDVEIQLD 340
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      341  SGRIYFDGGLKSTMLAHPKIDPETRELFALSYDVVSKPYLKYFRFTSDGKSPDVEIPLD 400

QY      341  QPTMHDFALTENFVVVPDQVQVFKLPMIRGGSPVVDKMKARFGLTOKIYAEDSSNIK 400
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      401  QPTMHDFALTENFVVVPDQVQVFLRPMIRGGSPVVDENKKSPFGILNNAKDASSIQ 460

QY      401  WIDAPPCFCFLHNAWEPPETDEVVIGSCMTPPDISIFNESDENLKSVLSEIRNLKTGE 460
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      461  WIEVPDPCFCHLWNSWEEPETDEVVIGSCMTPPDISIFNEHDETLQSVLSEIRNLKTGE 520

QY      461  STRRPISNEDQVNLNLEAGWNRMLGRKTFAYLALAEPPKPKYSGFAKVDLTTEGVKKH 520
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      521  STRRPVIS---EGVNLNLEAGWNRMLGRKTRVAYLALTEPPKPKYSGFAKVDLSIGTERKY 577
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      521  LYGNRYGGPELPLPGEGGDEBGTILCFVHDEKTKWSELOLVNNAVSLEVEATVKLPSRY 580
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

[illegible]

```

QY 521 LYCDNRYGSEPLPLPGEGGDEBGGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSPRV 580
DB 536 MYGEERFGEPLFLP-NGOKEDDGYILAFVHDEKWKSELOIVNAQNKLKLEASIKLPSPRV 594
QY 581 PYGFHGTFTIGADDLAKQ 597
DB 595 PYGFHGTFTIHSKDLRKQ 611

RESULT 7
QY 582 YGFHGTFTIGADDLAKQ 597
DB 599 YGFHGTFTIHSKDLRKQ 614

ID Q9M6E8 PRELIMINARY; PRT; 615 AA.
AC Q9M6E8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED1.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, TOP CROP;
RX MEDLINE=20079657; PubMed=10611388;
RA Qln X., Zeevaert J.A.;
RT "The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory
RT step of abscisic acid biosynthesis in water-stressed bean.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:15354-15361(1999).
DR EMBL: AF190462; AAF26356.1;
DR InterPro: IPR004294; RPE65.
DR Pfam: PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 615 AA; 68075 MW; OCC10F862D7DE130 CRC64;

Query Match 68.7%; Score 2165; DB 10; Length 615;
Best Local Similarity 69.2%; Pred. No. 1.7e-160;
Matches 426; Conservative 59; Mismatches 95; Indels 36; Gaps 9;

QY 11 SGRWLGNHTQPPPLSSQSSDLSCSS-----LPMASRVTRKLN-VSSALHIPPALHPKP 64
DB 6 SNTWI--NTTLPSSCSPPKDLASTSSPTLLFPKKRSSNTNTITCSLQI---LHPVK 60
QY 65 Q-----SSNSPALVWPKAKE-----SNTKQ-----MNLFORAAALDAAE 101
DB 61 QYQPTSTSTTTTPKPTTTTTTTPHRETCKPLSDTKQPPKQKNFLQKAAATGLDME 120
QY 102 GFLVSHKHLPLPKTADPSVOIAGNFAPVNEQPVRRNLVVYVKLPDSIKGYVVRNGANPL 161
DB 121 TALVSHKSHKPLPKTADPKVOIAGNFAPVPEHAADQALPVVGKTPKCIDGYVVRNGANPL 180
QY 162 HEPVTHGHHFDDGGMHVAHVEHSGSAYACRFTQTNRFVQROGLGRPVFPKAIKELHGT 221
DB 181 YEPVAGHHFDDGGMHVAHVEHSGSAYACRFTQTNRFVQROGLGRPVFPKAIKELHGS 240
QY 222 GIARLMLFYARAAGIIVDPAGCTGVANAGIYFNCRLILAMSEDDLPYQVQITPNDGLKTV 281
DB 241 GIARLLLFYARSLFQLVDGSHGGMVANAGIYFNNHLLAMSEDDLPYHVRITSNGLDTY 300
QY 282 GRFDGQLESTMLAHKVPDPESGFLFALSVDVYVKPKLYFRSPDGTGKSPDVEIQLDQ 341
DB 301 GRYDFNGQLNSTMLAHKLPDVPNGDLHALSYDVQKPKLYFRSADGVKSPDVEIPLKE 360
QY 342 PTMHDFAITENFVVPDQVVFVKLPENIRGSGSPVYVDKNKVARFGILDKYAEISSNIKW 401
DB 361 PTMHDFAITENFVVPDQVVFVKLTETMITGSGSPVYVDKNTSRGILDKNAKANAMRW 420
QY 402 IDAPDCCFHLNWAWEPEDETVVIGSCMTTPDISFNESENKLSVSEIRLNKTKGES.461
DB 421 IDAPDCCFHLNWAWEPEDETVVIGSCMTTPDISFNESENKLSVSEIRLNKTKGS 480

```

```

QY 462 TRPPIISNEDQVNLNLAGVYNNRNLGRKTKFAYLALAEPPKYSGGFAKVDLTITGEVKKHL 521
DB 481 TRPPIIS-DAEQVNLNLAGVYNNRNLGRKTKFAYLALAEPPKYSGGFAKVDLFSGEVOKYM 539
QY 522 YGDNRYGSEPLPLPGEGGDEBGGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSPRV 581
DB 540 YGEKFGGEPFLP-NGEEEGDGYILAFVHDEKWKSELOIVNAQNKLKLEASIKLPSPRV 598
QY 582 YGFHGTFTIGADDLAKQ 597
DB 599 YGFHGTFTIHSKDLRKQ 614

RESULT 8
QY 582 YGFHGTFTIGADDLAKQ 597
DB 599 YGFHGTFTIHSKDLRKQ 614

ID Q9AX24 PRELIMINARY; PRT; 625 AA.
AC Q9AX24;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED3.
OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, LULA;
RA Chernys J., Zeevaert J.A.D.;
RT "Abscisic acid in avocado fruit.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF224671; AAK00623.1;
DR InterPro: IPR000221; Protamine_pl.
DR InterPro: IPR004294; RPE65.
DR Pfam: PF03055; RPE65; 1.
DR PROSITE: PS00048; PROSITE.
KW Dioxygenase.
SQ SEQUENCE 625 AA; 69724 MW; A56F33042D5F2CB6 CRC64;

Query Match 65.6%; Score 2066.5; DB 10; Length 625;
Best Local Similarity 64.9%; Pred. No. 8.9e-153;
Matches 392; Conservative 74; Mismatches 93; Indels 45; Gaps 6;

QY 38 LPMASRVTRKLNVSALHTPP-----ALHFP-----KQSS 67
DB 22 LPISKNLRSRPNFTMLKHNTPLIQCCSHSPSSSSAAVHLHLPKQPTKSKPSIKKGSST 81
QY 68 NSPATVVKPKAKESNTKQ-----MNLFORAAALDAAEGLVSH--EKLHPLPKT 116
DB 82 LTPSIEKNPGSHOVKTDQSGPNRVGPNWNIQRTAAAFALDAIEKLIARVLERHRLPKT 141
QY 117 ADPSVQIAGNFAPVNEQPVRRNLVVYVKLPDSIKGYVVRNGANPLHEPVTGHHFDDGGM 176
DB 142 ADPEVQIAGNFAPVNEQPVRRNLVVYVKLPDSIKGYVVRNGANPLHEPVTGHHFDDGGM 201
QY 177 VHAYKFEHSGSAYACRFTQTNRFVQROGLGRPVFPKAIKELHGTGIARLMLFYARAAG 236
DB 202 IHAVKFRNGSASYCRFTQTNRFVQROGLGRPVFPKAIKELHGTGIARLMLFYARAAG 261
QY 237 IVDPAHGTGVANAGIYFNGRLLAMSEDDLPYQVQITPNDGLKTVGRFDGQLESTMLA 296
DB 262 LVNADEGMGVANAGIYFNGRLLAMSEDDLPYHVRITPNDGLKTVGRFDGQLESTMLA 321
QY 297 HPKVDPESEGLFALSVDVYVKPKLYFRSPDGTGKSPDVEIQLDQPTMMHDFAITENPV 356
DB 322 HPKLDPESEGLFALSVDVYVKPKLYFRSPDGTGKSPDVEIQLDQPTMMHDFAITENPV 381
QY 357 VPDQGVTFKLEMEIRGSGSPVYVDKNKVARFGILDKYAEISSNIKWIDAPDCCFHLNWA 416
DB 382 IPDQGVTFKLEMEIRGSGSPVYVDKNKVARFGILDKYAEISSNIKWIDAPDCCFHLNWA 441

```

```

QY 417 EBPETDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRIISNEDQOVNL 476
Db 442 EBPESGEVVVIGSCMTPPDSIFNESDENLKSILTEIRLNTRIGESTRRRII-DPQPLNL 500
QY 477 EAGMVRNRMGLRKTTFAYLALAEPPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPLFLPG 536
Db 501 EAGMVRNRMGLRKTTFAYLALAEPPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPLFLPG 560
QY 537 E---GGEDEGYILCFVHDEKWKSELOLVNVAISVEVATVILPSRVPYGFHGTFCAGDD 593
Db 561 EYSTSGREDDGYVVFHDEKTSRSELLILNANNRLEASVMLPSRVPYGFHGTFCISSRD 620
QY 594 LAKQ 597
Db 621 LAKQ 624

RESULT 9
Q9C6Z1
ID Q9C6Z1 PRELIMINARY; PRT: 589 AA.
AC Q9C6Z1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase, putative.
GN T2H7.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=41130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT *Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.;
RL Nature 408:816-820(2000).
DR EMBL; AC074176; AAG50855.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 589 AA; 65336 MW; 6A79B2BF7EAA8179 CRC64;

Query Match 64.9%; Score 2044.5; DB 10; Length 589;
Best Local Similarity 68.1%; Pred. No. 4.2e-151;
Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps 4;

QY 27 SSSLSYSCSSLPMSRVTRKLNLSALHTPPALHFFKSSNSPAIVVKKPKAKESNTKQM 86
Db 27 SPSSVSFTNTKPR---RRKLSANSVDTNLLNPNVPSNPPII-----PEKDTSRW 76
QY 87 NLFQRAAALDAARGFLVSHKHLPLKPTADPSVQIAGNFAPVNEQPVRRNLPPVVGKLP 146
Db 77 NLFQRAAALDAFAETALLRRERSKPLKPTVDRHQISGNVAPPEQSVKSSLSVDGKIP 136
QY 147 DSIKGVYVRNGANPLPEVPTGHHFFDGDGMVHAVKFEHGSASACRFTQTNTFVQERQLG 206

```

```

Db 137 DCIDGYLVRNGANPLPEVPSGHLEFDGDMVHAVKTTNGDASYSCTETETRLVQEKQLG 196
QY 207 RVFPAKAIAGELGHTGIARIMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLIAMSDDL 266
Db 197 SPIFPAKAIAGELGHSGIARIMLFYARGLFGLLNHNKNGTGVANAGLVYFHDRLAMSEDDL 256
QY 267 PYOVLTTPNGDLKTVGRFDFDQLESTMTAHKVDPPESGELFALSVDVYVKPKLYKFRFS 326
Db 257 PYOVRTDNGDLTIGRDFDQLSAMTAHKIDPVIKELFALSVDVYVKPKLYKFRFS 316
QY 327 PDGTSKPDVEIQDPTMMHDAITENFVVDPDQVQVFKLPEMIRGSGPVVYDKNKVARF 386
Db 317 PEGEKSPDVEIPLASPTMMHDAITENFVFDQVQVFKLSDFLGSKSPVKDGEKISRF 376
QY 387 GILDYAEDESSNIKWDAPDCFCFHLNWAWEPEPTEDEVVIGSCMTPPDSIFNESDENL 446
Db 377 GILPRNAKDAEMVWVESPEFCFHLNWAWEPEPTEDEVVIGSCMTPADSIENECDEQLN 436
QY 447 SVLSEIRLNKLTGSTRRIISNEDQOVNLKAGMVRNRMGLRKTTFAYLALAEPPKVS 506
Db 437 SVLSEIRLNKLTGSTRRIIPG-SVQMLNLAGMVRNRLGRKTRAYLALAEPPKVS 495
QY 507 FAKVDLTGTEVKKHLYGDNRYGGEPLFLPG--EGGEDEGYILCFVHDEKWKSELOLVN 564
Db 496 FAKVDLTGTEVKKHLYGDNRYGGEPLFLPG--EGGEDEGYILCFVHDEKWKSELOLVN 555
QY 565 AVSLEVEATVILPSRVPYGFHGTFCAGDDLAKQ 597
Db 556 AVTLEATVILPSRVPYGFHGTFCAGDDLAKQ 588

RESULT 10
Q49505
ID Q49505 PRELIMINARY; PRT: 583 AA.
AC Q49505;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE NEOXANTHIN cleavage enzyme-like protein.
GN F28J12.10 OR AT4G18350.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021710; CAAL6715.1;
DR EMBL; AL161548; CAB78837.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
SQ SEQUENCE 583 AA; 65066 MW; A138F93542E50852 CRC64;

Query Match 63.2%; Score 1991; DB 10; Length 583;
Best Local Similarity 64.5%; Pred. No. 6.3e-147;
Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;

```



```

QY 75 -RPAKESNTKOMNLPQRAAAALDA-ABGFLYS-BEKLHPLKTDPSVQIAGNEAPVN 131
Db 72 AAPKAEAGGKQQLNLFQRAAAALDAFEEGFVANVLERPHGLSTADPAVQIAGNEAPVG 131
QY 132 EQPVRENLPWVKLPDSIKGVYVRNGANLHEPVTGHHFFDGMVHAKVFEHGA-SYA 190
Db 132 ERFPVHELPSVGRIPPFIDGVYARNGANFCFDPVAGHHFLFDGMVHALKIRNGAAESYA 191
QY 191 CRTQTNRRFVQERQOLRPVFPKPAIGELHGHGTGIALMLFYARAAAGIIVPAHGTGVANAG 250
Db 192 CRTETARLQERARIGRPVFPKPAIGELHGHGTGIALMLFYARAAAGIIVPAHGTGVANAG 251
QY 251 LVYFNGRLAMSEDDLPYQVQITPNDLKTGVRGDFDGMVHAKVFEHGA-SYA 310
Db 252 LVYFNGRLAMSEDDLPYQVQITPNDLKTGVRGDFDGMVHAKVFEHGA-SYA 311
QY 311 SYDVVSKPYLKPFRFPGDGTGSDVLEQPTMIHDAITENFVVPDQVVKLPPEMI 370
Db 312 SYDVIRKPYLKPFRFPGDGTGSDVLEQPTMIHDAITENFVVPDQVVKLPPEMI 371
QY 371 RGSPVYVYDKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNAAWEEPEDETVVYVIGSC 430
Db 372 RGSPVYVYDKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNAAWEEPEDETVVYVIGSC 431
QY 431 MPPDSIFNESDENLKSIVSEIRLNKLTGSTRRPISNEDQOVNLEAGMVRNMLGRKT 490
Db 432 MPPDSIFNESDENLKSIVSEIRLNKLTGSTRRPISNEDQOVNLEAGMVRNMLGRKT 490
QY 491 KPAYLALAPWPKVSGFAVDLTTGVKKHLYGDNRYGGEPLFLPGEGE-DEEGY 545
Db 491 KPAYLALAPWPKVSGFAVDLTTGVKKHLYGDNRYGGEPLFLPGEGE-DEEGY 545
QY 546 ILCFVHDEKTSKELQIYNAVSLVEATVKLPSRVPYGFHGTFIGADDLAKQ 597
Db 551 VLFVHDERAGRSLLVWNAADIRLEATVQLPSRVDFGFGHTFIGQELEAQ 602

RESULT 13
Q9LRM7 PRELIMINARY; PRT: 577 AA.
AC Q9LRM7;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 9-cls-epoxycarotenoid Glucose-6-phosphate
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty pl and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB028621; BAB01363.1;
DR InterPro; IPR004294; RPF65.
DR Pfam; PF03055; RPF65; 1.
KW Dioxigenase.
SQ SEQUENCE 577 AA; 63821 MW; 515E0A1B321B1FE1 CRC64;

```

Query Match 52.8%; Score 1663.5; DB 10; Length 577;
 Best Local Similarity 55.4%; Pred. No. 2.4e-121;
 Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

```

QY 24 LSSQSSDLSYCSLPMASVTRKLVSS-ALHTFPALHPKQSSNSPAIVVVKPAKESN 82
Db 9 LLPTKTSRSHLLPQPKANISRILNPKIPTLPDLTSPVSP-----VKLKPTYPN 62
QY 83 TKOMNLFQRAAAALDAABE-GELVSHKELHPLKTDPSVQIAGNEAPVNEQVRENLPV 141
Db 63 ---LNLQKLAATMLDKIESSIVPMQONRPLPKDPAVQLSGNFAPVNECEVQNGLEV 119
QY 142 VGKLPDSIKGVYVRNGANLHEPVTGHHFFDGMVHAKV--PEHGSASYACRFTOTNR 199
Db 120 VGQIPSLKGVYVRNGANLHEPVTGHHFFDGMVHAKV--PEHGSASYACRFTOTNR 178
QY 200 VQERQOLRPVFPKPAIGELHGHGTGIALMLFYARAAAGIIVPAHGTGVANAGLVYFNGRL 259
Db 179 VOETALGRSVFPKPAIGELHGHGTGIALMLFYARAAAGIIVPAHGTGVANAGLVYFNGRL 238
QY 260 AMSEDDLPYQVQITPNDLKTGVRGDFDGMVHAKV--PEHGSASYACRFTOTNR 319
Db 239 AMSEDDLPYQVQITPNDLKTGVRGDFDGMVHAKV--PEHGSASYACRFTOTNR 298
QY 320 LKTFRSPDGTGSDVLEQPTMIHDAITENFVVPDQVVKLPPEMI 379
Db 299 LKTFRSPDGTGSDVLEQPTMIHDAITENFVVPDQVVKLPPEMI 358
QY 380 KKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNAAWEEPEDETVVYVIGSC 435
Db 359 KKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNAAWEEPEDETVVYVIGSC 417
QY 436 SIPNESDENLKSIVSEIRLNKLTGSTRRPISNEDQOVNLEAGMVRNMLGRKT 495
Db 418 SIPNESDENLKSIVSEIRLNKLTGSTRRPISNEDQOVNLEAGMVRNMLGRKT 473
QY 496 ALAEPKPYSGFAVDLTTGVKKHLYGDNRYGGEPLFLPGEGE-DEEGY 545
Db 474 ALAEPKPYSGFAVDLTTGVKKHLYGDNRYGGEPLFLPGEGE-DEEGY 533
QY 556 WKSELIQIYNAVSLVEATVKLPSRVPYGFHGTFIGADDLAKQ 598
Db 534 DESEFVVDATDMKQVAAVRLPERVYGFHGTFIGADDLAKQ 576

RESULT 14
O49675 PRELIMINARY; PRT: 595 AA.
AC O49675;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NEOXANTHIN cleavage enzyme-like protein (AT4G19170/T18B16_140).
GN T18B16.140 OR AT4G19170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bervan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Bervan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Ban J.,

```

